Dieter Gertmann

WIRKUNG DES VOLLZUGS DER AKADEMISCHEN PRÜFUNG AUF SELBSTBEWERTUNG UND MOTIVATION DER PRÜFLINGE

1977

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41547213_0005

Psychologie

Wirkung des Vollzugs der akademischen Prüfung auf Selbstbewertung und Motivation der Prüflinge

Inaugural-Dissertation
zur Erlangung des Doktor- Grades
der
P h i l o s o p h i s c h e n F a k u l t ä t
der
Westfälischen Wilhelms-Universität
zu
Münster (Westf.)

vorgelegt von
Dieter Gertmann
aus Freiburg (Breisgau)
1977

Tag der mündlichen Prüfung: 17.2.1977

Dekan: Prof. Dr. L. Budde

Referent: Prof. Dr. M. Sader

Korreferent: Prof. Dr. W. Keil

UNIV.
LEIDEN
BIBL.

Ich danke allen, die geduldig gewartet haben.

Ich danke Herrn Professor Dr. M. Sader für die Betreuung der Arbeit.

Ich danke den Kollegen an der Pädagogischen Hochschule Münster für die Freisetzung von Verpflichtungen.

Ich danke den Mitarbeitern des Forschungsprojekts "Empirische Untersuchung der akademischen Prüfung": Helmut Alves, Gerd Gompert, Joachim Lodders, Bernhard Raming, Elisabeth Schliehe, Antje Seeger, Gisela Voth-Ollefs, Andrea Weingarten, Werner Wothke, Hermann Timmerbrink.

Ich danke Bernhard Raming für Hilfen bei der Datenverarbeitung und Klaus Herbst für Anregungen zur Auswertung der Nutzenskalierung.

Ich danke den Mitgliedern des "Doktorandenkolloquiums": Beate Clemens-Lodde, Bernd Jötten, Heike Keil-Specht, Dietrich Kost, Hans-Jörg Ulm.

Ich danke meiner Mutter für die große Sorgfalt beim Schreiben der Arbeit.

Inhaltsverzeichnis

		Seite
I.	Einführung in die Themenstellung	1
I.1.	Erläuterung der Themenstellung	1
I.2.	Einordnung der Arbeit	2
I.3.	Besonderheiten der Darstellung	2
II.	Hochschulpädagogische Vorüberlegungen und Hintergründe	3
	Übersicht	3
II.1.	Hochschuldidaktischer Rahmen der Untersuchung	4
	Einordnung und Begründung des hochschuldidaktischen Ansatzes	5
	Rolle der Psychologie	13
	Zusammenfassung	16
II.2.	Funktionen akademischer Prüfungen	17
	Historisch-soziologische Funktionentheorie der Prüfung	18
	Die Selektionsfunktion der Prüfung	20
	Die Sozialisationsfunktion der Prüfung	24
	Prüfung als Funktionenmischung	26
	Die Steuerungsfunktion der Prüfung	28
	Die Diagnosefunktion der Prüfung	34
	Zusammenfassung	39
II.3.	Rückmeldefunktion der Prüfung	40
	Struktur, Gegenstand und Inhalt der Rückmeldung	40
	Normwerte der Selbstbewertung	44
	Normwerte der Motivation	46
	Zusammenfassung	48

Seite

II.4. Leistungsbewertung und Leistung 49

Leistungsbewertung durch Noten 49
Leistungsbewertung und Leistungsprinzip 54
Zusammenfassung 59

III. Psychologische Analyse der Rückmeldefunktion 60

Übersicht 60
Bedeutung von Theorie und Empirie 62
Methodisches Vorgehen der empirischen Untersuchung 65

III.I. Voraussetzungsanalyse 68

Übersicht 68

III.I.A. Prüfungen als Hintergrund und Ursache 70

Prüfungsbedingungen am Psychologischen Institut der Universität Münster 70
Ebenen der Auswertung für die Fächer der mündlichen Prüfungen 75
Eingrenzung der Wirkung der Prüfungen 76
Zusammenfassung 78

III.I.B. Eigenschaften der Noten 80

Übersicht 80

III.I.B.1. Statistische und skalometrische Eigenschaften der Noten 81

Struktur der Notenskala (Prüfer) 81
Verteilungseigenschaften der Noten 86
Bildung von Gesamtnoten 90
Transformation der Noten 95
Zusammenfassung 96

III.I.B.2. Konstitution der Noten in Prüfungen 98

Zuverlässigkeit von Noten 99
Nachweis einzelner Fehlerquellen 102
Inhalte von Noten 104
Zusammenfassung 106

Seite

III.I.B.3. Konstitution der Noten dieser Untersuchung 107

Einfluß des Prüfers 107

Situative und personale Bedingungen der Einzelnoten 110

Situative und personale Bedingungen der Gesamtnoten 113

Zusammenhänge zwischen den Einzelnoten 113

Zusammenfassung 116

III.I.B.4. Eignung der Noten zur Rückmeldung 117

III.I.C. Selbstbewertung: Theorie und Methode 120

Vielfalt von Selbst-Begriffen 120

Selbstkonzept als Grundbegriff 124

Struktur und Prozeß beim Selbstkonzept eigener Leistungsfähigkeit 126

Methode der empirischen Erfassung 129

Validität des Konstrukts im Spiegel empirischer Befunde 132

Zusammenfassung 140

III.I.D. Motivation: Theorie und Methode 141

Grundcharakter der Motivation 142

Ableitung und Begründung der Fragen nach Beliebtheit und Handlungsplanung 144

Beiträge zur Validität der Motivationsfragen 145

Motivation vor der Prüfung 150

Zusammenfassung 153

III.I.E. Empirische Eigenschaften von Selbstkonzept und Motivation 154

Übersicht 155

III.I.E.1. Analyse der Antwortverteilungen zum Selbstkonzept 155

Zusammenfassung 160

Seite

III.I.E.2. Persönlichkeits- und Situationsbedingungen 161

Unterschiede zwischen den Prüfungsarten (VD und HD) 161

Unterschiede zwischen den "Fach-Prüfer-Kombinationen" 164

Gewählte Prüfungsordnung und Geschlecht 172

Zusammenfassung 173

III.I.E.3. Kontingenzen der thematischen Fragen 175

Kontingenzen innerhalb der Fragen 175

Kontingenzen zwischen den Fragen 181

Zusammenfassung 184

III.II. Wirkungsanalyse 186

III.II.A. Theoretische Klärung möglicher Wirkungen ... 186

Übersicht 186

Eingrenzung der Wirkungsanalyse 187

III.II.A.1. Wirkungen auf das Selbstkonzept eigener Fähigkeit 190

Konstitution der Selbstkonzepte in der Sozialisation 190

Situativer Einfluß auf Selbstkonzepte 191

Einfluß diagnostischer Information 202

Einfluß von Noten 203

Situationsbezogene Integration der Ergebnisse 205

Zusammenfassung 208

III.II.A.2. Wirkungen auf die Motivation 210

Einfluß im Modell der Leistungsmotivation ... 210

Zur Rolle der intrinsischen Motivation 214

Einfluß von Erfolg und Mißerfolg, Lob und Tadel 216

Einfluß diagnostischer Information 217

Einfluß der Noten 219

Situationsbezogene Integration der Ergebnisse 221

Zusammenfassung 223

Seite

III.II.B. Empirische Analyse der Wirkungen 224

Übersicht 224

Methodische Grundlagen der Wirkungsanalyse .. 225

III.II.B.1. Globale Wirkung der Noten 230

Wirkung der Noten in den "Fach-Prüfer-Kombinationen" 230

Wirkung der Noten auf der Ebene "Über-alle-Fächer" 235

Wirkungen der Gesamtnoten auf die Gesamtkonzepte 241

Zusammenfassung 243

III.II.B.2. Geschlecht der Prüflinge und gewählte Prüfungsordnung (APO/NPO) als Rahmenbedingungen 245

Abhängigkeit der Wirkung von der Wahl der Prüfungsordnung 245

Abhängigkeit der Wirkung vom Geschlecht der Prüflinge 248

Zusammenfassung 250

III.II.B.3. "Fach-Prüfer-Kombinationen" des VD als Rahmenbedingungen 251

Zusammenfassung 255

III.II.B.4. Einfluß der Prüfung auf Kontingenzen zwischen den thematischen Fragen 256

Änderung der Korrelationen zwischen den thematischen Fragen 257

Änderung der Korrelationen innerhalb der thematischen Fragen 261

Zusammenfassung 263

III.II.B.5. Einfluß der Prüfung auf die Antwortverteilungen der thematischen Fragen 264

Zusammenfassung 267

Seite

III.II.B.6. Strukturelle Einbettung der thematischen Fragen 268

Anwendung der Faktorenanalyse als Methode der Strukturanalyse 268

Faktorenstruktur im VD 270

Faktorenstruktur im HD 274

Zusammenfassung 278

III.III. Stufenanalyse 280

Übersicht 280

Inhaltliche und methodische Struktur der Stufenanalyse 281

III.III.1. Subjektive Notenskalen der Prüflinge 283

Theorie, Methode und funktionale Bedeutung subjektiver Notenskalen 284

Empirische Ergebnisse der Auswertung der Notenskalen 289

Zusammenfassung 297

III.III.2. Das Anspruchsniveau als Grundlage erlebten Erfolgs und Mißerfolgs 299

Zusammenfassung 306

III.III.3. Die erlebte Selbstverantwortlichkeit für die Noten 306

Zusammenfassung 317

III.III.4. Die erlebte Rückmeldung in der Prüfung 318

Zusammenfassung 324

IV. Praktische Konsequenzen für die Rückmeldung im Studium 325

Zusammenfassung 333

V. Zusammenfassung der Arbeit 335

Seite

Exkurs I. Planung, Durchführung und Auswertung der empirischen Untersuchung ..... E I 1

Übersicht ..... E I 1

E I.1. Methodische Anlage der Untersuchung ..... E I 1

Erkundungscharakter der Untersuchung ..... E I 2

Verwendung der Methode schriftlicher Befragung ..... E I 3

Anlage der Untersuchung als prae-post-design E I 5

Form der Fragen ..... E I 6

E I.2. Planung und Durchführung der Untersuchung E I 11

Entwicklung der Fragebögen ..... E I 12

Endform der Fragebögen ..... E I 14

Gewinnung der Daten ..... E I 16

E I.3. Datenqualität und Datenauswertung ..... E I 19

Besonderheiten der Daten ..... E I 20

Datenverarbeitung ..... E I 21

Anwendung statistischer Verfahren ..... E I 21

Auswirkungen der Selbstselektion ..... E I 25

Formale Qualität der Daten ..... E I 29

Exkurs II. Auswertungsebenen der Antworten zu Einzelfächern ..... E II 1

Anhang I. Wortlaut der Fragen (PVF und PFF) ..... A I 1

A I.1. Thematische Fragen ..... A I 1

Prüfungsvorbereitungsfragebogen ..... A I 1

Prüfungsfolgenfragebogen ..... A I 3

A I.2. Übrige Fragen ..... A I 4

Prüfungsvorbereitungsfragebogen ..... A I 5

Prüfungsfolgenfragebogen ..... A I 8

L Literaturverzeichnis ..... L 1

I.: Einführung in die Themenstellung

I.1.: Erläuterung der Themenstellung

"Wir gehen an den Examina zugrunde; die meisten, welche sie bestehen, sind dann so abgewirtschaftet, daß sie irgendeiner Initiative unfähig sind, sich gegen alles, was an sie herankommt, möglichst ablehnend verhalten und, was das Schlimmste ist, eine große Meinung von ihren Fähigkeiten haben, weil sie siegreich aus allen diesen Examina herausgekommen sind."

Das auf dem Umweg über PAULSEN (1902) hierher wie zu PRAHL (1976) gelangte Zitat BISMARCKs stellt den Aspekt der akademischen Prüfung heraus, der in Absetzung von vielen anderen empirischen und theoretischen Arbeiten über Prüfungen diese Untersuchung beschäftigt: die Wirkung des Vollzuges der Prüfung auf die Prüflinge selbst. Die Wirkung wird in zwei möglichen Richtungen gesehen; zum einen als Wirkung der Prüfung auf die leistungsbezogenen Urteile der Prüflinge über sich selbst (Selbstbewertung), zum anderen als Wirkung der Prüfung auf die Handlungsbereitschaft (Motivation) der Prüflinge. Die beiden Begriffe Selbstbewertung und Motivation deuten an, daß die Analyse der Prüfung sich der begrifflich-theoretischen wie der methodischen Mittel bedienen will, die die Psychologie zur Verfügung stellt. Erstes Ziel ist weniger die Anreicherung psychologischen Fachwissens als vielmehr die Erweiterung hochschuldidaktischen Wissens über Prüfungen mit den Mitteln der Psychologie. (Es wird sich zeigen, daß beide Zielsetzungen sich nicht ausschließen müssen.)

Das Primat der Hochschuldidaktik fordert, das psychologische Vorgehen analytisch und normativ in die Ziele einzupassen, die mit dem Gegenstand auch dessen Untersuchung tragen. Es fordert zugleich, den Ergebnissen über Erkenntnisse hinausgreifend praktische

Folgerungen hinzuzufügen. Die folgend kurz dargestellte Aufteilung in die (ansonsten von Inhalt und Umfang kaum vergleichbaren) drei Hauptteile trägt diesen Forderungen Rechnung:

Teil II Hochschulpädagogische Vorüberlegungen

Teil III Ergebnisse der theoretischen und empirischen Untersuchung

Teil IV Praktische Konsequenzen

I.2.: Einordnung der Arbeit

Die vorliegende Arbeit ist im Rahmen eines Forschungsprojektes mit der Themenstellung "Empirische Untersuchung der akademischen Prüfung" entstanden, das von der Stiftung Volkswagenwerk in den Jahren 1970-72 gefördert wurde. Ergebnisse des Projektes sind außer in dieser Arbeit in mehreren Diplomarbeiten dargestellt (ALVES & LODDERS 1974; SCHLIEHE & SEEGER 1974; TIMMERBRINK 1971; WOTHKE 1974).

I.3.: Besonderheiten der Darstellung

Die Darstellung weist einige Strukturierungen auf, deren Begründung zu einer Annäherung der Wirkung an die Absicht beitragen mag.

1. Vor den Kapiteln (römische Zahlen) oder Teilen (Buchstaben, arabische Zahlen) der Arbeit wird in einer Übersicht der Verlauf der Darstellung angedeutet, um über allen Einzelheiten einen Zusammenhang zu stiften.

2. Die einzelnen Kapitel oder Teile werden jeweils von Zusammenfassungen abgeschlossen. Die Zusammenfassungen sollen jene Argumentationen und Ergebnisse herausstellen, die für den Fortgang der Darstellung oder eine direkte Verwendung über die Arbeit hinaus wesentlich scheinen.

3. Zwei ausführlichere Teile, die den Gang der Argumentation eher stören würden, sind im Anhang als Exkurse zusammengefaßt.

Exkurs I: Durchführung der empirischen Untersuchung
Exkurs II: Ebenen der Auswertung

Die Exkurse sind nur durch ihre Zusammenfassung im laufenden Text vertreten.

II.: Hochschulpädagogische Vorüberlegungen und Hintergründe

Übersicht: Der hochschuldidaktischen Zwecksetzung der Arbeit entspricht eine Einordnung des Ansatzes in die unterschiedlichen Richtungen der Hochschulforschung. Die konkreten Absichten des Ansatzes werden mit den hochschulpolitischen Gegebenheiten und Tendenzen konfrontiert. Die Grundlinien hochschuldidaktischer Forschung werden in inhaltlicher und methodischer Hinsicht bestimmt. In einem weiteren Schritt wird die allgemeine Bedeutung der Psychologie in dem gewählten Ansatz durch ihre positiven Funktionen und negativen Implikationen verdeutlicht. Das soll im ersten Teil des Kapitels, stichwortartig und mehr oder weniger verkürzt, angesprochen werden.

Prüfungen sind als didaktische Prozesse durch die Funktionen bestimmt, die sie in der Universität und im Wechselfeld von Universität und Gesellschaft haben. Historische und strukturell-systemhafte Betrachtungen dienen als Material einer Funktionsanalyse, die versucht, die soziologisch verstandenen Funktionen auf die psychologische Ebene der Betrachtung zu überführen. Das soll im zweiten Teil dieses Kapitels geschehen.

Als Thema der Arbeit wird aus der Funktionsanalyse heraus die Diagnosefunktion der Prüfung, sofern sie die Wirkung der Prüfung auf die einzelnen Prüflinge betrifft (Rückmeldefunktion), festgelegt. Es kann aufgezeigt werden, daß die theoretischen und empirischen Mittel der Psychologie sich gerade für eine kritische Analyse dieser Funktion eignen, wenn Motivation und Selbstbewertung als die wesentlichen Ziele der

Funktion gesehen werden. Die Analyse der Wirkung von Prüfungen auf Selbstbewertung und Motivation hat dabei ihren Ausgangspunkt in Zielbestimmungen zu nehmen, die vorab Normwerte für die Richtung und das Ausmaß des wünschbaren Einflusses formulieren. Dem gilt der dritte Teil dieses Kapitels.

Dieses so beabsichtigte Vorgehen macht in einem vierten Teil die Beschäftigung mit den u.a. für diese Funktion institutionell vorgesehenen Mitteln, den Noten, erforderlich. Dabei muß und kann wie bei der allgemeinen Funktionsanalyse auf die historische Betrachtungsweise zurückgegriffen werden. Gleichzeitig ist eine Auseinandersetzung mit der von Noten deutlich vorgenommenen Reduktion inhaltlich aspektreichen Verhaltens auf Leistungen notwendig.

II.1.: Hochschuldidaktischer Rahmen der Untersuchung

Da es den Ausführungen dieses Teils durch ihre Beschränkung auf Andeutungen an Schärfe in der Bestimmung der eigenen Position mangelt, sei sie vorab verdeutlicht.

Der gewählte Ansatz ist von zwei Seiten kritisierbar:

1. Er geht von einem gegenüber einem umfassenden Anspruch von Hochschuldidaktik auf die methodische Betrachtung reduzierten Verständnis aus.

2. Er ordnet sich in Reformansätze ein, die Antworten auf krisenhafte Entwicklungen im deutschen Hochschulwesen geben wollen. In der Zwischenzeit aber sind die Krisen administrativ und durch Teilreformen bewältigt oder durch Krisen in anderen Bereichen abgelöst, so daß die Verwirklichung von Reformen im Sinne der ursprünglichen Absicht erschwert wird.

Die eigene Position äußert sich in den Antworten auf diese beiden Kritikpunkte.

1. Die Beschränkung auf die partikulare Orien-

tierung an methodischen Aspekten ist aus dem Gegenstand Prüfung selbst begründet, der sich im wesentlichen in seiner Besonderheit als Methode erweist. Wegen seiner Partikularität erfordert das methodenorientierte Vorgehen um so dringender die konkrete Einbettung in Zielvorstellungen. Reformziel der Untersuchung ist der Abbau von Fremdbestimmung durch eine Vergrößerung des Ausmaßes an Selbstbestimmung.

2. Die für Reformansätze ungünstigere Ausgangssituation gilt in besonderem Maße für Prüfungen, denen eine verstärkte Bedeutung aus der administrativen Formierung des Studiums zuwächst. Demgegenüber sind zwei einander ergänzende Gegenstrategien sinnvoll: Auf der rational ausgerichteten Kritik an Prüfungen als Gegenüberstellung der behaupteten Funktionen mit wirklichen Leistungen der Prüfungen ist auch dann zu bestehen, wenn die Subjekte, die aus dieser Kritik Folgerungen ziehen könnten, nicht in Sicht sind (PRAHL 1976). (Das ist als der Versuch zu sehen, in der Theorie erfolgreich zu "überwintern".) Und gleichzeitig sind stellenweise die immer noch gegebenen Freiräume durch zielgerichtete partikulare (u.U. örtlich beschränkte) Reformen auszunutzen.

Einordnung und Begründung des hochschuldidaktischen Ansatzes

Die gegenwärtige Situation der Hochschulforschung an westdeutschen Universitäten kann kaum als der rationale Versuch reflektierter Selbststeuerung der Arbeits- und Bildungsprozesse im tertiären Bereich betrachtet werden (ECKSTEIN 1972). Sie ist mehr noch als im angloamerikanischen Bereich eine isolierte Reaktion auf verschiedenartige Konflikte innerhalb der Universitäten und zwischen Universität und Gesellschaft. Je nach Art der Konflikte und deren Sichtweise sind die Ansätze unterschiedlich. Sie verstehen sich

auch keinesfalls selbst immer als einem einheitlichen Forschungsgebiet zugehörig. Die unterschiedlichen Ansätze innerhalb des weiten Rahmens sollen aufzählend genannt werden, wobei die Reihenfolge der einzelnen Ansätze keine sachliche Priorität sondern eher eine historische Anordnung verdeutlichen soll.

1. Es finden sich Ansätze der Bildungsökonomie, die in einem deskriptiv vergleichenden bildungsstatistischen Vorgehen besteht (z.B. EDDING 1962) und mittelbar die praktische Forderung nach quantitativer Ausweitung des tertiären Bereichs aus den vergleichenden Bildungsstatistiken begründet. Hier wird der Konflikt zwischen einer an den Bildungsbegriffen des 19. Jahrhunderts orientierten elitären Ideologie der Hochschulen und den nachgearbeiteten Anforderungen der Industriegesellschaft des 20. Jahrhunderts deutlich. In diesem Zusammenhang ist die Ausweitung der Studentenzahlen von ungefähr 30 000 im Jahre 1900 (PRAHL 1976) auf 200 000 bis 250 000 zu Beginn der 60er Jahre und auf eine Zahl, die sich im Augenblick der Millionengrenze nähert, zu erwähnen.

2. Ist der Konflikt zwischen Anspruch auf Elitenbildung und gesellschaftlicher Notwendigkeit erweiterter Bildungschancen in dem Ansatz der Bildungsökonomie kaum auf eine inhaltlich wertende Ebene der Betrachtung gehoben, so leistet dies wegweisend der Band "Hochschule in der Demokratie" (NITSCH et al. 1965), der eine Denkschrift des SDS aufnimmt und an der kritischen Theorie orientiert das Verhältnis von Hochschule und Gesellschaft zu bestimmen versucht. Ziel ist eine den gegenwärtigen gesellschaftlichen Widersprüchen angepaßte strukturelle Reform der Universität, um deren Autonomie wiederherzustellen und gesellschaftlich wirksam werden zu lassen.

3. Einen im engeren Sinn empirischen Ansatz repräsentieren die Studienerfolgsuntersuchungen. Dabei

haben sie auch die sozialen (KATH et al. 1966) oder nur die individuellen Voraussetzungen (GASCH 1971) des Abschlusses eines Studiums im Blick. In der auf individuelle Selektion gerichteten Ausprägung entsprechen sie einem umfangreichen Zweig der amerikanischen Hochschulforschung in Form der prediction-Untersuchungen (LAVIN 1965; SCHUMANN & CLAUS 1970). Sie sind allesamt ein Mittel einer anders verstandenen Bildungsökonomie, der es um den, gemessen an institutionellen oder individuellen Erfolgskriterien, rationellen Einsatz der Bildungsmittel des tertiären Bereichs geht, wobei die Rationalisierung nicht allein in der Selektion gesucht werden muß (BEARD et al. 1964).

4. Eine Intensivierung erfährt die Beschäftigung mit der Hochschule im Verlauf der als Studentenbewegung aufkommenden Jugendbewegung Mitte bis Ende der 60er Jahre. Anknüpfend an die Auseinandersetzung um den gesellschaftlichen Stellenwert der tertiären Bildung (s. 2.) wird kämpferisch eine Umwandlung universitärer Machtstrukturen gefordert, nicht zuletzt in der Absicht, damit gleichzeitig eine gesellschaftliche Wandlung zu fördern (LEIBFRIED 1967). Hier setzt auch massivere Kritik an Prüfungen ein (MÜLLER 1967).

5. Verunsichert durch den studentischen Protest und die Entwicklung der Hochschulen zum Massenbetrieb, nimmt mit Ende der 60er Jahre die Bereitschaft zur reformorientierten Beschäftigung mit den Formen der tertiären Bildungsprozesse zu. Gefördert wird dies durch die Bundesassistentenkonferenz und den Verband Deutscher Studentenschaften (SPINDLER 1968). Bedeutsam sind auch die Aktivitäten des "Arbeitskreises für Hochschuldidaktik" mit seinen Schriftenreihen "Blickpunkt Hochschuldidaktik" und "Hochschuldidaktische Materialien". Die Beschäftigung mit den Lern- und Arbeitsprozessen im tertiären Bereich versucht, die Gunst der Verunsicherung für die Reform der herkömmlichen Formen des Lehrens und Lernens zu nutzen, um

die Ausbildung wissenschaftlicher, reflektierter, kooperativer und herrschaftsfreier zu machen (s. dazu den Zielkatalog der BAK, Dokument 11 in: SPINDLER 1968). Die Ausformulierung dieser Ziele zu methodisch orientierten Vorschlägen findet unter maßgeblicher Beteiligung von Psychologen statt (SADER et al. 1970), die sich an einer im angloamerikanischen Bereich bestehenden Tradition empirischer Forschung (und zeitgemäßerer Universitätsausbildung) orientieren können.

Auch die vorliegende Arbeit geht vom Versuch aus, mit den Mitteln der Psychologie zu einer methodischen Reform beizutragen. Gegenstand ist die Prüfung, deren Kritik und Reform von Anfang an wesentlicher Bestandteil der reformorientierten Ansätze ist (SCHÜTZ et al. 1969).

Diese Hochschuldidaktik im engeren Sinn integriert weder die ihr vorangestellten Ansätze noch u.a. die folgenden:

- die von einem kritischen Standpunkt ausgehende Hochschulpsychiatrie (DÖRNER 1967), die schon vorher mit Prüfungen sich auseinandersetzte;
- den Versuch der Pädagogik, die Einheit von Forschung und Lehre als Einheit von Wissenschaftsdidaktik und Hochschuldidaktik neu zu begründen (v. HENTIG 1972).

Die darin deutlich werdende "Beschränktheit" der Hochschuldidaktik (die man eher als Hochschulmethodik bezeichnen sollte; siehe den englischen Titel des Werkes "Lehren und Lernen an der Hochschule": "Research into Teaching Methods in Higher Education" BEARD 1968, 1972) fordert eine explizite Begründung in drei Aspekten: erstens in der bedingenden faktischen Möglichkeit von Reformen; zweitens in der Berechtigung, mit dem Gegenstand der Methoden des Lehrens und Lernens die konkreten Inhalte ausklammern zu können; drittens in der Wahl des Vorgehens wissenschaftlicher Erkenntnisgewinnung.

Die Bereitschaft der Universität und der sie tragenden politischen Strukturen, die in ihr üblichen Formen der Lehre durch die Hochschuldidaktik in Frage stellen zu lassen und neue Formen zu erproben, läßt sich, wie erwähnt, auf das Zusammentreffen der quantitativen Ausweitung der Universitäten mit der Studentenbewegung zurückführen, die beide selbst nicht unabhängig voneinander zu sehen sind. Inzwischen sind begrenzte Reformen erfolgt, obwohl diese besonders in der Reform des Studiums hinter den einmal gesetzten Zielen weit zurückbleiben. Das nicht erwartet beschleunigte Anwachsen der Studentenzahlen aber hat, zusammen mit der verschlechterten finanziellen Situation und anderen gesellschaftlich-politischen Faktoren, den Freiraum eingeengt, der für hochschuldidaktische Reformen auch notwendig ist. Es sind Entwicklungstendenzen im Hochschulsystem der Bundesrepublik absehbar, die aus dem Zwang zu rationellerer Nutzung vorhandener Mittel das Studium in starre Zeitvorgaben pressen (Hochschulrahmengesetz: Regelstudienzeiten). Ein weiteres Moment der Abwendung von Reformen ist die Verlagerung der Konflikte an die "Randbereiche des Studiums": Konfliktbeladen ist zum einen die Zulassung zum Studium mit dem immer noch massiv wirksamen Numerus clausus; zum anderen zeichnen sich Konflikte im Bereich der Berufszugänge ab, die sowohl die politische Kontrolle als auch das zumindest sektorale Überangebot an Absolventen des tertiären Bereiches betreffen. Alle drei Entwicklungstendenzen - Rationalisierung, Zugang und Berufsselektion - funktionalisieren gerade auch Prüfungen verstärkt und engen so über das allgemeine Klima hinausgehend ganz konkret die mögliche Reform der Prüfung ein.

War das Interesse an Hochschuldidaktik und der Einsatz für hochschuldidaktische Reformen auf der Seite der Lehrenden immer schon auf eine kleine Gruppe von Hochschullehrern beschränkt, ist ein Interesse auf

seiten der Lernenden mit dem Abklingen der Studentenbewegung wieder weitgehend erloschen, so tendiert der zunehmende administrative Druck auf die Hochschulen zu Lösungen, die mehr nur der Verwaltungsrationalität verpflichtet sind. Doch darf dies nicht daran hindern, die gerade auf diesem Gebiet vernachlässigte theoretische und empirische Gegenstandsaufklärung weiterzutreiben. Auch sollte nicht übersehen werden, daß mehr als in anderen Bereichen weiterhin Freiräume auszunutzen sind, um Konsequenzen der Gegenstandsaufklärung zu verwirklichen.

Zur wissenschaftlichen Legitimierung aber ist auch dann, wenn die praktische Absicht als reformerische einen Ansatz auszeichnet, eine zusätzliche Begründung von Inhalt und Methode notwendig. Die inhaltliche Begründung setzt an dem Begriff der Didaktik an. (Im folgenden wird der Begriff der Methode in zwei Bedeutungen verwendet: Methode als Gegenstand der Forschung und Methode als Vorgehen der Forschung. Hier meint "Methode" erst einmal in Absetzung von "Inhalt" die formale Charakteristik der Lehr- und Lernprozesse.) Es wurde schon angedeutet, daß der Begriff Didaktik bei der Beschränkung auf Methoden des Lehrens und Lernens als Gegenstand verkürzt verwendet wird. In der Allgemeinen Pädagogik hat sich der Begriff der Didaktik als sehr vielfältig erwiesen (ROTH 1976), doch ist allen Konzeptionen eigen, daß die Didaktik eine Bestimmung der Lerninhalte wesentlich (mit)umfaßt (BLANKERTZ 1969).

Genau diese Bestimmung der Inhalte bleibt in einem Vorgehen, das sich allein mit der formalen Charakteristik der Prozesse des Lehrens und Lernens, beispielsweise einer Vorlesung, einer Übung oder eben auch einer Prüfung beschäftigt, ausgespart. Die Aussparung ist fragwürdig, weil zwischen Inhalten und Methoden ein Implikationszusammenhang (BLANKERTZ 1969) angenommen werden muß. Der Implikationszusammenhang

kann jedoch sowohl von der Seite des Inhaltlichen als auch von der Seite des Methodischen her aufgesprengt werden. Eine prinzipielle Priorität der Inhalte besteht nicht (MENCK & THOMA 1972). Es erscheint deshalb als durchaus angemessen, bei bestimmten Gegenständen didaktischer Forschung den Methodenaspekt zu verabsolutieren. Prüfungen legen das nahe, wie die ansonsten auf Integration von Methode und Inhalten angelegte Schrift "Forschendes Lernen - Wissenschaftliches Prüfen" der BAK (1970) zeigt.

Doch ist damit noch nicht die generelle Beschränkung hochschuldidaktischer Forschung auf das Methodische zu rechtfertigen. Sie ermöglicht zwar die einheitliche Warte einer empirisch orientierten Sozialwissenschaft. Ein Gesamtkonzept ohne Integration der Inhalte aber wäre funktional und normativ falsch.

Daraus folgt für diese ausschließlich methodisch orientierte Untersuchung: Ihr notwendiger Legitimationsgrund wäre ein didaktisches Gesamtkonzept. Das ist nicht in Sicht, u.a. weil es von der an die Psychologie angelehnten Hochschuldidaktik systematisch verfehlt wird. Deshalb ist das methodenorientierte Vorgehen vom besonderen Gegenstand her zu begründen. Darüber hinaus sind, weil ein integrierendes Konzept (Bildungsbegriff) fehlt, explizit die methodischen Prozesse normativ abzusichern, da sonst die Gefahr rein technischer Rationalisierung als dem methodenorientierten Vorgehen immanentes Ziel besteht. Erst die normative Absicherung belegt auch den Anspruch auf Reformabsichten.

Als dritter Aspekt muß das methodische Vorgehen im Sinne der Entscheidung für eine bestimmte Forschungsmethode bedacht werden. (Auch hier gilt ein Implikationszusammenhang zwischen wissenschaftlichem Gegenstand und wissenschaftlicher Methode, den die Erörterung nur implizit einbezieht.)

Die Hochschuldidaktik versteht sich als empirische Sozialwissenschaft. Ihre Methode ist eine Mischung

aus der Übertragung allgemeiner sozialwissenschaftlicher Erkenntnis auf den spezifischen Gegenstand und aus eigenständigen empirischen Untersuchungen. Formal lassen sich ihre Ansätze wie solche zur Bewertung von Bildungsprozessen allgemein unterteilen in strukturbezogene, prozeßbezogene und produktbezogene (HAGEN & THORNDIKE 1960; KEIL & PIONTKOWSKI 1973). Die Entwicklung der empirischen Lehrforschung zeigt dabei eine Tendenz weg von den produktbezogenen Ansätzen, den "Kriterien des Erfolgs"-Modellen (GAGE 1967), hin zu den prozeßbezogenen Ansätzen. Auch in der Hochschuldidaktik sind prozeßbezogene Paradigmen angebracht. Die Entwicklung ergibt sich aus der Erfolglosigkeit der produktbezogenen Ansätze. Zugleich aber entsprechen die prozeßbezogenen und strukturbezogenen Ansätze eher den Reformabsichten.

Doch sind mit der Entscheidung gegen produktbezogene Ansätze noch nicht alle methodischen Grundprobleme gelöst. Ein weiteres z.B. schließt sich in Form des Objekt-Subjekt-Verhältnisses in sozialwissenschaftlicher Forschung an: Sozialwissenschaftliche Forschung tendiert dazu, die Personen oder Organisationen als Gegenstände der Untersuchung zu Objekten zu machen. Personen und Organisationen aber sind zugleich Subjekte der Prozesse und Strukturen und als solche anzusprechen, wenn Reformen demokratisch über Aufklärung erreicht werden sollen. Die Gefahr der empirischen Sozialwissenschaften besteht darin, den Zweck der Aufklärung zweifach durch ihre Methoden zu verfehlen, indem Methoden angewendet werden, die für die Betroffenen nicht mehr einsehbare Erkenntnisse ergeben; und indem mit den Methoden, die vorgeblich neutral scheinen, bestimmte Ziele eingeführt werden, die wegen dieser scheinbaren Neutralität unbestimmt bleiben.

Einen Versuch, durch aktive Beteiligung der Personen und Organisationen am Forschungsprozeß selbst die beiden Gefahren auszuschließen, stellt die Aktionsforschung dar (HAAG et al. 1972; GROEBEN &

WESTMEYER 1975), die gerade auch für den Bereich der Hochschulforschung geeignet ist (SANFORD 1970). Sie soll durch die Hereinnahme des Subjektcharakters der Objekte in einem mehrstufigen Vorgehen die Subjekt-Objekt-Trennung aufheben. Die Aktionsforschung ist aber noch mehr Programm und in der Durchführung von vielen Voraussetzungen abhängig, wie u.a. einer relativ homogenen Reformabsicht der Beteiligten. Wenn diese Untersuchung deshalb auf der Subjekt-Objekt-Ungleichung beharrt, sind deren negative Folgen nur dadurch zu mildern, daß die Ergebnisse auf die angewendeten Methoden rückbezogen werden. Das müßte bei den "Abnehmern" der Ergebnisse eine eigenständige Zielklärung begleiten, so daß die Untersuchung nur einen ersten Schritt eines Diskussionsprozesses darstellen kann.

Rolle der Psychologie

Die Einschränkung auf Methoden als Gegenstand der Hochschuldidaktik und die Betonung des prozeßbezogenen Vorgehens machen die feststellbar enge Anbindung der Hochschuldidaktik an Erkenntnisse und Methoden der Psychologie verständlich und sinnvoll. Dem positiven Beitrag der Psychologie stehen aber auch Nachteile gegenüber, die bestimmte Eigentümlichkeiten der Psychologie als Wissenschaft über die schon dargestellten methodischen hinausgehend in die Hochschuldidaktik hineintragen.

Der positive Beitrag der Psychologie besteht nach WEINERT (1967 in Anlehnung an HEITGER) in den drei Aufgaben:der Realitätskontrolle des pädagogischen Handelns, der Grundlage für die Setzung neuer Ziele, der Ideologiekritik pädagogischer Normen durch Erforschung der Wirklichkeit. Etwas abgewandelt lassen sich dem die Begriffe Konstruktion, Zielfindung und Analyse zuordnen.

Die Psychologie kann zur Zielfindung im pädagogischen Bereich nur in einem beschränkten Umfang beitragen. Nach BOKELMANN (1965) lassen sich Normen pädagogischen Handelns in verschiedene Klassen unterteilen. Die übergeordneten Sinnormen für die Erziehung werden durch die Psychologie kaum tangiert; in der diesen nachgeordneten Klasse der Normen der Erziehung finden sich didaktische, asketische und methodische Normen. Die Psychologie stellt begriffliche und methodische Mittel bereit, um diese Normen (insbesondere die methodischen) zu explizieren und zu konkretisieren, was als Bestandteil des Zielfindungsprozesses gesehen werden muß. Von diesem Angebot der Psychologie soll in der anschließenden Ausformulierung der Ziele und Funktionen von Prüfungen Gebrauch gemacht werden.

Im Rahmen der Analyse pädagogischer Prozesse kommt der Psychologie die Rolle kontrollierter Erfahrungsbildung zu. Als solche ist sie auch in dieser Untersuchung zentral, indem ihre empirischen und begrifflich-theoretischen Mittel zur Überprüfung von aus Zielformulierungen abgeleiteten Forderungen an Verlauf und Ergebnis von Prüfungen als pädagogischen Prozessen eingesetzt werden.

Die konstruktive Bedeutung der Psychologie liegt im Überschneidungsbereich der beiden ersten Aufgaben, wo für Ziele Möglichkeiten zu deren Verwirklichung gefunden werden sollen. Dabei können die Ziele im Prozeß ihrer Verwirklichung selbst eine Veränderung erfahren.

Die Untersuchung macht von allen drei Einsatzmöglichkeiten der Psychologie in der dargestellten Reihenfolge Zielfindung, Analyse und Konstruktion Gebrauch. (In dem erwähnten Modell der Aktionsforschung ist dieser Ablauf über die Verwirklichung von konstruktiven Vorschlägen vermittelt mehrfach und abgewandelt zu durchlaufen. Das bleibt, wie erwähnt, der aktiven Rezeption der Ergebnisse überlassen.)

Dem positiven Einsatz der Psychologie steht nun eine Reihe von Einwänden gegenüber, die gegen die (isolierte) Anwendung der begrifflichen und methodischen Mittel der Psychologie erhoben werden können. Die folgende Aufzählung ist nur ein erster allgemeiner Hinweis, der jeweils an Ort und Stelle bestimmter Begriffe und Methoden Beachtung finden müßte. Da das die Darstellung überfordert, erfolgt die Kritik vorab.

1. Mit HOLZKAMP (1972) ist die Unfähigkeit der gegenwärtigen akademischen Psychologie zur historischen Betrachtung zu betonen. Die Psychologie stellt Prozesse als naturgegeben hin, die sicher nur in einer bestimmten historischen Situation oder, noch eingegrenzter, in einer bestimmten sozialen Umgebung oder sozialen Gruppe in einer historischen Situation Gültigkeit haben. Sie übersieht dabei die nicht personalen Bedingungen des menschlichen Verhaltens und Erlebens, deren Beeinflussung scheinbar individuelle Probleme universeller lösen hilft.

2. Die Psychologie ist entsprechend eher geneigt, die Kausalität für bestimmte Vorgänge in den Personen zu suchen als in den organisatorischen und sozialen Bedingungen, unter denen die Personen handeln. So weist DÖRNER (1967) bei "pathologischen" Reaktionen auf Prüfungen darauf hin, daß die Reaktionen nicht als Pathologie der sie zeigenden Personen gesehen werden müssen sondern als Folge der Pathologie der (Prüfungs-)Bedingungen, unter denen Personen sinnvoll so reagieren.

3. Die Psychologie reduziert einschneidende existentielle Ereignisse (wie z.B. Lebenskrisen) auf technisch neutralisierte Variablen: Beispielsweise wird die Angst, die in bestimmten Situationen so intensiv auftreten kann, daß sie lebensentscheidende Bedeutung hat, durch die Technik ihrer Messung (Operationalisierung) auf einen "gewöhnlichen" Prozeß reduziert. Die von MOELLER (1967) in Fallstudien in dem

psychoanalytischen Modell erklärte Angst in Prüfungen ist kaum noch mit der Prüfungsängstlichkeit zu vergleichen, die als alltägliche in Graden des Mehr oder Weniger mit Hilfe eines Fragebogens erfaßt wird (BROWN & GELDER 1938). Bei der letzteren kann der Verdacht auf eine die Struktur der Person beeinträchtigende Wirkung überhaupt nicht mehr aufkommen. Ein Aspekt dieser Reduktion ist die (unberechtigte) Gleichsetzung von Prozessen ganz unterschiedlicher zeitlicher Erstreckungsgrade (KAMINSKI 1964), wenn an experimentell manipulierten Prozessen kurzer Dauer gewonnene Erkenntnisse über die Verwendung gleicher Begriffe auf zeitlich langdauernde Lebensprozesse übertragen werden. Ein anderer Aspekt betrifft die voreilige Bereitschaft zur Verwendung von eingeführten Methoden ohne Prüfung, ob sie der Spezifität der Situation angemessen sind. Auf der operationalen Basis wird damit die Begriffsübertragung vorentschieden.

Die ersten beiden Kritikpunkte sollen in der folgenden Darstellung insofern berücksichtigt werden, als die strukturellen Bedingungen der Prüfung einer psychologischen Differenzierung der prozessualen Abläufe vorgeordnet betrachtet werden.

Zusammenfassung (II.1.)

Der Ansatz einer reformorientierten Hochschuldidaktik, zu der sich diese Untersuchung rechnet, wird in dreierlei Hinsicht konkretisiert:

1. Die Reformorientierung wird mit den gegen Reform gerichteten Tendenzen im Hochschulwesen konfrontiert. Gegenüber diesen Tendenzen soll an einer Reformorientierung festgehalten werden, die die immer noch gegebenen Freiräume nutzt.

2. Es wird aufgezeigt, daß die Hochschuldidaktik einen auf methodische Aspekte eingeschränkten Didaktikbegriff verwendet. Das kann im Fall der Prüfung durch

den Gegenstand selbst legitimiert werden. Es macht aber zusätzlich eine besonders deutliche normative Begründung des Vorgehens erforderlich, um nicht der Illusion einer inhaltlich neutralen Technik aufzusitzen.

3. Das Vorgehen der Untersuchung setzt sich als prozeß- und strukturbezogen vom produktbezogenen Paradigma der Bildungsforschung ab.

Die Bedeutung der Psychologie wird getragen vom methodischen und prozeßbezogenen Charakter des Ansatzes. Positiv kann die Rolle der Psychologie als Beitrag zur Zielfindung, als Analyse und Beitrag zur Konstruktion von Alternativen (Reform) gesehen werden. Gefahren aus der Anwendung der Psychologie ergeben sich aus deren unhistorischer und personenbezogener Sichtweise sowie aus der unkritischen Übertragung von Begriffen und Methoden auf Gegenstände in ganz unterschiedlichen Kontexten und von deutlich andersartigem Grundcharakter (wie beispielsweise der zeitlichen Erstreckung).

II.2.: Funktionen akademischer Prüfungen

Prüfungen stehen als Methoden in einem Funktionszusammenhang. Der Funktionszusammenhang kann auf zwei verschiedenen Ebenen bestimmt werden, die verkürzt durch die beiden wissenschaftlichen Begriffsysteme der Soziologie und der Psychologie ausgefüllt werden. Die historisch-soziologische Betrachtung der Stellung der Prüfungen im Spannungsfeld von Hochschule und Gesellschaft geht strukturbezogen vor. Sie bildet den Rahmen, in den sich die psychologische, eher prozeßbezogene Betrachtung einordnet, ohne jedoch an diesen Rahmen in allen Aspekten gebunden zu sein. Beide Betrachtungsweisen liefern Erkenntnisse im Hinblick auf die normative Absicherung der Prüfung, die explizit erst für die eingegrenzte Rückmeldefunktion als Thema der Arbeit versucht werden soll; sie ergeben analytische Einsichten in die Strukturbedingungen und Prozeßabläufe; als

solche Einsichten binden sie auch konstruktive Überlegungen zu zielbezogenen Änderungen von Prüfungen. Dabei dient im folgenden Teil wie in den übrigen dieses Kapitels der Einsatz der Psychologie in erster Linie der Zielfindung und der Konkretisierung des Hintergrundes der Rückmeldefunktion.

Historisch-soziologische Funktionentheorie der Prüfung

Vereinfacht gesehen können Funktionsbetrachtungen in zwei Richtungen weisen: Sie können Funktionen in einem normativen Sinn postulieren, oder sie können Funktionen in einem analytischen Sinn konstatieren. Die historisch-soziologische Betrachtung ist in einem analytischen Sinne gemeint. Diese Festsetzung jedoch gerät in Widerspruch zur dargestellten Notwendigkeit eines normativen Rahmens, was noch dadurch verschärft wird, daß der analytisch-funktionalen Betrachtung der Vorwurf gemacht werden kann, von sich aus eine normative Bedeutung zu gewinnen. Die Lösung des Widerspruchs ist in den Ausführungen dieses Kapitels enthalten: Die Gesamtbetrachtung der Prüfungsfunktionen stellt nur den Hintergrund dar für die Untersuchung der Wirkung des Vollzugs der Prüfung auf den Prüfling selbst. Die analytische Betrachtung des Hintergrundes dieser Teilfunktion in Form einer Zusammenstellung aller Funktionen ist gerade deshalb ausreichend, weil die Teilfunktionen in unterschiedliche Richtungen weisen. Eine zielbezogene Reform ist deshalb nicht in einer einheitlichen Änderung der Prüfung zu suchen, sondern nur in einer zielbezogenen Differenzierung. Für die übrigen Funktionen reicht dann die analytische Betrachtung aus, um entscheiden zu können, ob eine zielbezogene Reform der Teilfunktion noch im Rahmen der Prüfung als einheitlicher Institution möglich ist oder der Rahmen aufgegeben werden muß. (Dabei ist vorausgesetzt, daß für die Teilfunktion

positiv bewertete Ziele als Grundlage der Reform herausgearbeitet werden können.)

Wenn in der Überschrift die Betrachtung als historisch-soziologische gekennzeichnet wird, ist beabsichtigt, die historische Betrachtungsweise als Hilfsmittel zur Aufdeckung der gegenwärtigen sozialen Funktionen von Prüfungen einzusetzen. Die geschichtlichen Wandlungen von Prüfungenkönnen in der Parallelität zu gesellschaftlichen Wandlungen gleiche Funktionen deutlich werden lassen. Solche Wandlungen können aber auch die Funktionen selbst betreffen; der gegenwärtige Zustand zeichnet sich dann erst auf dem Hintergrund geschichtlicher Funktionsverschiebungen ab. Ein dritter Ertrag der historischen Betrachtung besteht darin, daß die Ähnlichkeit von Prüfungen in unterschiedlichen sozialen Systemen die strukturelle Verknüpfung von Prüfung und Gesellschaft hervortreten läßt (etwa im Vergleich des Prüfungssystems des chinesischen Kaiserreiches mit dem der europäischen Universitäten).

In diesem Sinn ist eine Geschichte der Prüfungen gebunden an eine Sozialgeschichte der Universität, deren Erarbeitung noch weitgehend aussteht (PRAHL 1976).

Die selektiven europäischen Universitätssysteme und die diesen zugeordneten Prüfungen zeigen im großen und ganzen eine über nationale Besonderheiten hinweg einheitliche geschichtliche Entwicklung. Das amerikanische System dagegen ist in seiner mehr auf Allgemeinbildung gerichteten, weniger selektiven Struktur abweichend, doch sind in den letzten Jahrzehnten Konvergenzen festzustellen (FRAENKEL 1967; KANDEL 1936). Die folgenden Darstellungen gründen sich insbesondere auf die Arbeiten von PRAHL (1976) und KVALE (1972). Die Anlehnung an die beiden Arbeiten ist enger, als durch einzelne Autorennennungen angezeigt wird.

Die soziologische Analyse hat vom zentralen Stellenwert der Prüfung in der Vermittlung zwischen Hochschulsystem und Gesellschaft auszugehen (KVALE 1972). Im Rahmen der Vermittlung zwischen Universität und Gesellschaft fungieren die Prüfungen als Selektionsinstrumente und als Sozialisations/Normierungsinstrumente. Vereinfacht könnte man formulieren: Die Universität wirkt über die Selektionsfunktion der Prüfungen auf die Gesellschaft ein, während die Gesellschaft über die Sozialisations/Normierungsfunktion auf die Universität Einfluß nimmt. Jedoch erscheinen die Einwirkungsprozesse in einer differenzierten Analyse als wesentlich komplexer und zugleich widerspruchsvoller.

Grundlegend für die Bestimmung der Wirkungen innerhalb der beiden Funktionen sind Annahmen über das Ausmaß der gesellschaftlichen Autonomie von Universitäten und ihrer Prüfungen, die sehr unterschiedlich ausfallen. Während BOURDIEU & PASSERON (1971) und PRAHL (1976) von einer relativen Autonomie ausgehen (so auch die Hochschuldenkschrift von NITSCH et al. 1965), wird auf der anderen Seite eine Autonomie völlig bestritten (KVALE 1972). Doch scheint die Möglichkeit historischer Sedimentierungen (PRAHL 1976) die Annahme relativer Autonomie gerade bei Prüfungen zu rechtfertigen. Auch KVALE (1972) teilt in der konkreten Analyse implizit diesen Standpunkt, wenn er etwa feudalistische Elemente in den gegenwärtigen Prüfungssystemen zu erkennen glaubt. Selbst dort, wo die gegenwärtigen Prüfungen als reines Mittel des Staates erscheinen, setzt die geforderte Prüfungsreform (VDS 1973) die Autonomie zumindest im Prinzip voraus.

Die Selektionsfunktion der Prüfung

Die begriffliche Bestimmung der Funktion ist einfach. Die Selektionsfunktion (Auslesefunktion) meint die Zuerkennung der Berechtigung zur Einnahme einer be-

stimmten sozialen, meist beruflichen Position in Abhängigkeit vom Ergebnis der Prüfung. Mit der Zuerkennung der Berechtigung ist an den europäischen Universitäten im Gegensatz zu den USA von seiten der Studenten die Erwartung verbunden, eine Berufsposition bestimmter Höhe mit dem Bestehen der Prüfung besetzen zu können (KANDEL 1936). Die Sicherheit ist in vielen Fällen objektiv begründet. Aber die ungeplante Abstimmung der Ausbildung und der Beschäftigungschancen kann auch Disparitäten zur Folge haben, so im Deutschland zwischen den beiden Weltkriegen und bei der im Augenblick zu erwartenden Lehrerarbeitslosigkeit; noch deutlicher in Schweden oder in der Juristenausbildung an italienischen Universitäten.

Eine historische Ableitung der Selektionsfunktion geben LAUWERYS & SCANLON (1969) und PRAHL (1976). Funktional ähnliche Vorläufer selektiver Universitätsprüfungen sind die Beamtenprüfungen des chinesischen Kaiserreiches. Während der Entwicklung der kontinentalen Universitäten im ausgehenden Mittelalter hat die Selektionsfunktion eher eine untergeordnete Bedeutung. Sie dient im wesentlichen der Selbstrekrutierung, indem sie Formen der Selektion aus anderen Ständen übernimmt (u.a. aus dem Handwerk; HARTOG 1938), was der Universität hilft, den gegen sie aufkeimenden Verdacht von Kirche und Gesellschaft zu besänftigen. Im Laufe der Zeit wird die Ablegung von (Staats-)Prüfungen dann als Erfordernis für die Berufstätigkeit in juristischen, ärztlichen und kirchlichen Berufen verlangt. Zur Zeit des Absolutismus ist mit dem Verfall der Universitäten ein gleichzeitiger Verfall des Prüfungssystems festzustellen. Die universitären Prüfungen werden zunehmend kommerzialisiert, indem die akademischen Grade käuflich erwerbbar werden; Staatsprüfungen als Zugangsvoraussetzungen allerdings bleiben unverändert. Nach der bürgerlichen Revolution wird die Auslese für Tätigkeiten in der staatlichen Bürokratie, zunehmend auch

die in nichtstaatlichen Professionen durch Prüfungen reguliert, die am Konkurrenzprinzip orientiert der Ablösung feudalistischer Vorrechte dienen sollen (LOCH 1963). In England wird beispielsweise eine Kommission eingesetzt, die die Erfahrungen des chinesischen Beamtenprüfungssystems aufarbeiten soll. Die Konkurrenzprüfungen, die eine am Leistungsprinzip orientierte Demokratisierung gegen den Adel verwirklichen, sind gleichzeitig politisch regressiv, indem sie die über Bildung vermittelte Beteiligung des "4. Standes" an gesellschaftlicher Macht behindern (BEHREND 1929).

In der Entwicklung der Auslesefunktion in den letzten Jahrzehnten zeichnen sich die folgenden Tendenzen ab: Fachwissen gewinnt eine zunehmende Bedeutung gegenüber Bildungswissen (FRAENKEL 1967). Die Kontrolle der universitätsinternen Prüfungen durch externe Prüfer nimmt zu (LAUWERYS & SCANLON 1969). Eine ähnliche Tendenz in Deutschland bildet schon seit Mitte des vorigen Jahrhunderts die zu beobachtende Zurückdrängung von akademischen Prüfungen durch Staatsprüfungen (PAULSEN 1902). Auch Zwischenprüfungen werden entgegen mancher Absichten bei ihrer Einführung in den Rahmen der Auslesefunktion eingespannt (WENKE 1967).

Zusammenfassend ergibt sich eine geschichtlich zunehmende Bedeutung der Selektionsfunktion. Kurzfristig wird die Tendenz noch dadurch unterstrichen, daß der zu erwartende Absolventenüberschuß in vielen Studiengängen, wie sich im Bereich der Lehrerausbildung zeigt, in Orientierung an den Prüfungsergebnissen kanalisiert werden wird.

Gilt diese Kennzeichnung bisher für die europäischen Universitätssysteme, gewinnen auch in den USA die Hochschulprüfungen mehr und mehr eine selektive Bedeutung (FRAENKEL 1967).

Die Abschätzung der gegenwärtigen Bedeutung nimmt,

wie beabsichtigt, keine Wertung der Funktion als solcher vor, sondern gibt nur eine Zusammenstellung von Hinweisen auf die Erfüllung der Funktion.

Als negative Bilanz schreiben BOURDIEU & PASSERON (1971) der Auslesefunktion das Entstehen einer hierarchischen Mentalität im Hochschulbereich zu und betonen, daß die angestrebte Demokratisierung durch die Prüfungen nicht erreicht werde (s.oben BEHREND 1929). Durch die Dokumentation formal gleicher Behandlung verschleiern Prüfungen soziale Selektionen in der Hochschule, die in Form von Studienabbruch (SCHALLBERGER 1974) schon vor den Prüfungen stattfinden. Auch die Prüfungen selbst sind nicht als sozial neutral zu betrachten; trotz schon weitgehender sozialer Selektion ist der Prüfungserfolg wiederum von der sozialen Herkunft abhängig (RUSAM 1972). Zudem verhindern die Prüfungen die Anerkennung von beruflich erworbenen Qualifikationen (CAPELLE in: LAUWERYS & SCANLON 1969).

Eine positive Leistung der Auslesefunktion ist die Sicherung von sozialer Stabilität (LAUWERYS & SCANLON 1969). Das wird beispielhaft an der in sehr unterschiedlichen sozialen Systemen gleichartig durch Ausleseprüfungen regulierten Zulassung zum Arztberuf deutlich (CHARVAT et al. 1968). Ein weiterer positiver Aspekt insbesondere der internen Prüfungen, aber auch des akademischen Einflusses in Staatsprüfungen, ist die Möglichkeit der Universität, über die Auswahl auf die gesellschaftliche Praxis einzuwirken (WENKE 1967). Eine Auslese durch abnehmereigene Prüfungen würde die relative Neutralität einer leistungsbezogenen Auslese durch ein Mehr an "Gesinnungskontrolle" ersetzen. Diese Gefahr gilt nur für den staatlichen und den halbstaatlichen Bereich, wo in vielen Fällen aber eine zweite Prüfung die Wirkung der universitätsgebundenen ersten Prüfung nochmals einschränken kann. Im Bereich privater Abnehmervon Absolventen ist die Orientierung an Prüfungen von vornherein gering (REESE 1970).

Die mögliche positive Wirkung der Auslese wird wesentlich bestimmt durch die Güte der Vorhersage beruflicher Leistung. Da das Problem an anderer Stelle ausführlich erörtert wird (Kapitel III.I.), muß hier der Hinweis genügen, daß die Güte der Prüfung in dieser Hinsicht sehr in Frage gestellt ist.

Die Sozialisationsfunktion der Prüfung

Die begriffliche Bestimmung ist weniger einfach als die der Selektionsfunktion. Bei PRAHL (1976) wird fast ausschließlich der Begriff der Initiation verwendet, der aber nur eine, in der mittelalterlichen Universität dominante Form der Sozialisation meint, die zwar in den gegenwärtigen Prüfungen noch nachklingt, aber nur einen Randaspekt bildet. Sozialisation und Normierung umfassen Einflüsse der Prüfung auf die Institution Universität und die in ihr arbeitenden Personen. Der Einfluß richtet sich auf organisatorische Strukturen, auf das Verhalten und auf die Inhalte. Er geht über die Normierung des Verhaltens in der Prüfungssituation selbst (LAUTMANN 1971) weit hinaus. Die Normen sind nur teilweise durch die Bestimmungen der Prüfungsordnung explizit vorgegeben; sie werden wesentlich mitbegründet durch institutioneneigene Traditionen und die individuelle Auslegung auf Grund der Macht des Prüfers (die besonders die Wirkung auf die Studenten betrifft). Für die Studenten als Prüflinge bedeutet die Sozialisation die Gewöhnung an eine bestimmte Form des Arbeitsverhaltens und den Zwang zur Beschäftigung mit bestimmten Inhalten in der Vorbereitung auf die Prüfung. Die Prüfung als Sozialisation ist ferner als der Versuch zu sehen, den Übergang zwischen zwei Rollensystemen, der Rolle des Studenten und der Rolle des in beruflicher Verantwortung Arbeitenden, durch Verhaltensvorschriften zu normieren und Konflikte in dieser Übergangssituation zu vermeiden. Der letzte Aspekt wird von PRAHL (1976)

besonders herausgestellt und in Anlehnung an die soziologische Theorie der Initiationsriten behandelt. Das, was an anderer Stelle als pädagogische Funktion von Prüfungen begriffen wird (DOHSE 1967; KVALE 1972), ist gleichfalls soziologisch als Sozialisation und Normierung zu verbuchen.

Am Beginn der geschichtlichen Entwicklung der Funktion steht deutlich der Aspekt der Initiation, den die Prüfung in Anlehnung an ständische Formen der Sozialisation von Übergängen zwischen Rollensystemen aufnimmt. Erst mehr auf die sekundären Systeme beschränkt entwickelt sich die Prüfung als Kontrolle der Leistungsfähigkeit der Ausbildungsinstitutionen durch staatliche oder kirchliche Gewalten (KANDEL 1936). Die Kontrolle greift mit der Ausweitung der Staatsprüfungen und des Systems externer Prüfer auch normierend in die universitäre Ausbildung ein.

Die neuere Entwicklung läßt die folgenden Tendenzen beobachten: Der Staat nimmt den normierenden Einfluß über die Genehmigung der Prüfungsordnungen strenger wahr (Suprematie des Staates; WENKE 1967). Prüfungen werden zunehmend ins Studium verlagert (Zwischenprüfungen, studienbegleitende Prüfungen) und beeinflussen umfassend das Studium. Ein humanisiertes und rationalisiertes Prüfungssystem stellt sich vermehrt in den Dienst der ökonomischen Ausbildung angeforderter Arbeitskräfte (KVALE 1972).

Die gegenwärtige Einschätzung der Sozialisations/Normierungsfunktion muß zwischen einem internen und einem externen Aspekt trennen. Unter dem externen Aspekt hat man die Gefahr des Verlustes der Hochschulautonomie zu sehen, wenn die Hochschulen in den Zwang von außen auferlegter Prüfungsordnungen geraten. Unter dem internen Aspekt besteht das Problem darin, daß die Funktion zur "hochschuldidaktischen" Scheinlösung der Unfähigkeit der Universitäten wird, ein vernünftig strukturier-

tes Studienangebot bereitzustellen.

Prüfung als Funktionenmischung

Abschließend ist festzuhalten, daß sich durch die Gleichzeitigkeit der beiden Funktionen Konflikte ergeben müssen. Die Auslesefunktion der Prüfung gerät in Widerspruch zur Sozialisationsfunktion, zumindest wenn man deren internen Aspekt gesellschaftlich sinnvoll zu bestimmen versucht (SALZMANN 1966; WALL 1958). So muß die Prüfung soziologisch als nicht widerspruchsfreie Einheit von zwei Funktionen, deren Bedeutung jeweils im Zunehmen begriffen ist, gesehen werden.

Die summarische Feststellung wird noch verstärkt durch die Analyse der Auffächerungen der beiden Funktionen, wie sie von PRAHL (1976) oder OPPENHEIM et al. (1966/1967) beschrieben wird. Die Auffächerung in verschiedene Teilfunktionen, die beispielsweise aus der Zuschneidung der Normierungsfunktion und der Auslesefunktion auf unterschiedliche Adressaten entsteht, und zudem durch die Ansammlung nur historisch verständlicher Sedimentierungen (wie z.B. ritueller Momente) mitbedingt ist, bringt eine Reihe von neueren einzelnen Widersprüchen mit sich. Deshalb kann die Prüfung auch als ein Komplex vieler einzelner Funktionen betrachtet werden, die unabhängig voneinander zu teilweise unterschiedlichen Optimierungen tendieren würden. (Das wird für viele Reformansätze zum Problem, die nur jeweils Teilfunktionen herausgreifen wollen.) Diesen Eindruck einer unübersehbaren Vielfalt versucht KVALE (1972) zu korrigieren. Er führt sowohl die beiden genannten Hauptfunktionen wie die ihnen zugeordneten Teilfunktionen auf eine einzige Einflußquelle zurück: auf die im Sinne der bestehenden Struktur der Gesellschaft ausgeübte Herrschaft durch die Prüfung. Er ordnet den groben Kennzeichnungen der Gesellschaftsstruktur jeweils Typen von Prüfungen zu, so dem Feudalismus die

rituelle Prüfung, dem Konkurrenzkapitalismus die bürokratische Prüfung, dem Monopolkapitalismus die rationalisierte Prüfung. Wenngleich Macht und Herrschaft durchaus als vereinheitlichendes Merkmal beider Hauptfunktionen anerkannt werden können, so überzeugt die schematische Zuordnung der Prüfungen nicht. KVALE selbst betont das Weiterbestehen ritueller Momente in den Prüfungen an westdeutschen Hochschulen, ihm bereiten zudem die bürokratischen wie rationalisierten Strukturen der Prüfungen in sozialistischen Ländern Probleme bei der Einordnung in sein Schema.

Doch sollte soviel aus der Analyse beachtet werden, was auch durch Beobachter aus andersgearteten politischen Positionen geteilt wird: Prüfungen nehmen in ihrer Bedeutung eher zu, weil sie zentral zwischen Ausbildungssystem und Gesellschaft vermitteln. Diese Bedeutungszunahme ist in gesellschaftliche Entwicklungen eingebettet, denen gegenüber die Universität selbst dann machtlos wäre, wenn in ihr Kräfte auf einen Abbau der Prüfung hin tendierten. Ein solches Interesse aber ist aus verschiedenen Gründen nicht in Sicht (COX 1966/67a; PRAHL 1976). Die Bedeutungszunahme steht im Zeichen einer Verschärfung der Auslesefunktion und einer eher außenorientierten Normierung. Die Universität kann versuchen, die äußeren Einflüsse durch eigene Ziele zu begrenzen oder durch weitgehende Selbstbestimmung der inhaltlichen Ausfüllung zu überformen. Das bedeutet aber gleichzeitig, daß in der Konzentration darauf weitere Aspekte der Funktionen von Prüfungen zurücktreten müssen. Andere Funktionen und Teilfunktionen finden besonders dann, wenn sie für die Universität allein von interner Bedeutung sind, eine optimalere Verwirklichung wahrscheinlich unabhängig von der Prüfung.

Das Bild von der Vielfalt und Widersprüchlichkeit der funktionalen Aspekte von Prüfungen bei gleich-

zeitiger Überformung durch die beiden Hauptfunktionen verstärkt sich noch, wenn zusätzlich zu Universität und Gesellschaft das Individuum als Determinante eingeführt wird. Dem individuumzentrierten Aspekt der Prüfung soll durch eine Transformation der beiden Hauptfunktionen auf die psychologische Betrachtungsebene mehr Aufmerksamkeit geschenkt werden. Die Transformation auf die psychologische Betrachtungsebene geht von den folgenden grob begrifflich angedeuteten Zuordnungen aus: Die Sozialisations/Normierungsfunktion der Prüfung wird als Steuerungsfunktion abgehandelt, die Auslesefunktion der Prüfung als Diagnosefunktion.

Die Transformation auf die Ebene von personalen Prozessen und Wechselwirkungen von Person und Situation bringt gegenüber der soziologischen Betrachtung eine Einschränkung und Erweiterung mit sich. Die Einschränkung betrifft den Umfang der Funktionen; er wird reduziert, indem institutionelle Gegebenheiten nur noch eine untergeordnete Rolle spielen. Deshalb ist die psychologische Betrachtung nur im Rahmen der soziologischen sinnvoll. Zugleich bewirkt die psychologische Betrachtung aber auch eine Erweiterung, indem sie in dem für das Gesamt der Prüfungsfunktionen nicht unwesentlichen Ausschnitt von individuellem Verhalten und Erleben eine konkrete Zielfindung und Analyse gestattet.

Die Analyse wird in diesem Teil vorgenommen im Hinblick auf die Steuerungsfunktion, weil diese eine gegenüber dem vorangehenden Teil weitere Präzisierung des funktionalen Hintergrundes erlaubt, auf den die Diagnosefunktion aufbaut, die dann anschließend zielbezogen bestimmt werden soll.

Die Steuerungsfunktion der Prüfung

Der Begriff der Steuerungsfunktion von Prüfungen ist in der Diskussion relativ neu. Er wurde von

SADER (1968) als "Prüfung als Studiensteuerung" geprägt. Er findet sich seitdem in der bundesrepublikanischen Diskussion um akademische Prüfungen häufig wieder (KVALE 1972; PRAHL 1976). KVALE, der die Steuerungsfunktion unter die pädagogische Funktion von Prüfungen subsummiert, deutet damit an, daß es nicht um jeden, wie der rein formale Begriff der Steuerung anzudeuten scheint, sondern um einen ganz bestimmten Einfluß der Prüfung auf das Verhalten der Studenten geht: die zielbezogene Wirkung der Prüfung auf Verhalten und Erleben der Studenten als Bewirkung von Lernprozessen und -inhalten. Diese Fragestellung wird in einer großen Zahl von begrifflichen Betrachtungen, in einer geringeren Zahl von empirischen Untersuchungen angesprochen.

Eine psychologisch adäquate Abhandlung der Steuerungsfunktion hätte eine Prozeßbetrachtung folgender Art anzustellen: Auf der Ebene des Verhaltens und Erlebens der Studenten wären die Prozesse zu kennzeichnen, die auf die Prüfung ursächlich zurückzuführen sind. Die Verursachung selbst wäre als Prozeß sozialer Beeinflussung zu erfassen, in den die Situationsdeutungen der Prüflinge, die sozialen Interaktionen zwischen den Prüflingen, aber auch Umweltbedingungen wie die Struktur der Ausbildungsinstitutionen vermittelnd eingreifen. Da nur eine empirisch wie theoretisch umfassende Studie der Steuerungsfunktion für eine spezifisch amerikanische Prüfung (Zulassung zum Doktorat) dieser Art vorliegt (MECHANIC 1962), die übrigen Untersuchungen die Prozesse auf einseitig gerichtete Einwirkungen reduzieren, ist die Darstellung auf die Unterteilung in abhängige und unabhängige Variablen beschränkt.

Während der überwiegende Teil der begrifflichen Analysen auf seiten der unabhängigen Variablen (Prüfungen) nicht mehr differenziert, beschäftigen sich die empirischen Untersuchungen im wesentlichen mit der Wirkung verschiedenartiger Prüfungsformen. In beiden Fällen

sind auf amerikanische Verhältnisse zugeschnittene Untersuchungen wegen der unterschiedlichen Rekrutierung der Studenten, der Dominanz studienbegleitender Prüfungen und anderer Strukturmerkmale des Hochschulsystems in ihren Ergebnissen nur mit Vorsicht zu übertragen.

Auf seiten der abhängigen Variablen stehen im Vordergrund kognitive Strukturbildungen und deren individueller Erwerb (konkret: Lernverhalten und Lernprodukte). Sowohl bei dem Lernverhalten als auch bei den Strukturbildungen kann zwischen quantitativen und qualitativen Einwirkungen der Prüfungen unterschieden werden.

Als quantitative Wirkung auf das Lernverhalten wird konstatiert: Prüfungen dienen als Anregung zur Arbeit, als Motivierung des Lernverhaltens (AGER & WELTMAN 1966/67; CHARVAT et al. 1968; EBEL 1958; HIMMELWEIT 1966/67; PREISS 1954/55; LAUWERYS & SCANLON 1969; SCHEER & ZENZ 1973). Sie veranlassen durch Zielkonkretisierung (COX 1966/67a; KVALE 1972) zu zielgerichtetem Lernen. Diese Wirkung wird implizit als mehr oder weniger positiv im Sinne einer Legitimierung gesehen.

Eher negativ bewertet steht dem die Ablenkung (CHARVAT et al. 1968), die Überarbeitung und Überreizung (PAULSEN 1902) durch die Prüfung entgegen. An dieser Stelle sind auch die das Arbeitsverhalten grundlegend störenden Auswirkungen der Prüfungsangst in ihren pathologischen Formen zu nennen (für viele andere: MOELLER 1967).

Die Ergebnisse spärlicher empirischer Untersuchungen können kaum eine Klärung bewirken. Es sind physiologische und psychologische Aktivierungskorrelate während der Prüfungsvorbereitung nachgewiesen (FISCH 1970; PIEHL 1972), denen konflikthaft hemmende Bestandteile beigemischt sind. Fallstudien zur Prüfungsangst belegen die mögliche aktivitätshemmende Wirkung der Prüfung (MOELLER 1967).

Als qualitative Wirkung auf das Lernverhalten wird positiv behauptet die Uniformität des Verhaltens durch gleiche Anforderungen (OPPENHEIM et al. 1966/67) sowie die Ausrichtung an wesentlichen Leistungen (Ebel 1958).

Demgegenüber ist eine umfangreiche Liste von negativ eingeschätzten Auswirkungen aufzustellen: Prüfungen verhindern unabhängiges und selbständiges Lernen (CHARVAT et al. 1968; JOHNSON & ABRAHAMSON 1968). Das angeregte Verhalten ist individualistisch (SCHEER & ZENZ 1973), flüchtig (JONES 1964/65) und extrinsisch motiviert (PRAHL 1976; PREISS 1954/55); häufigere Prüfungen führen zu einem Zerfall des Studiums (KURSITZA 1967; MARRIS 1964). Anstelle des sinnvollen Zieles, eine Berufsqualifikation zu erwerben, tritt die Ausrichtung des Verhaltens auf das Bestehen der Prüfung hin (SCHEER & ZENZ 1973).

Auch hier ist die Lage der empirischen Erforschung eher unbefriedigend. Die Feldstudie von MECHANIC (1962) weist eine große Zahl von Bewältigungsmechanismen (coping mechanisms) auf, die weder mit den Erwartungen der Prüfer noch mit den Idealen wissenschaftlichen Verhaltens in Übereinstimmung zu bringen sind. Unter anderem wird das individualistische Arbeitsverhalten als eine mögliche Bewältigungsstrategie beschrieben. ABERCROMBIE (1964/65) berichtet von der durch die Vorbereitung zur Prüfung verursachten Beendigung von sachorientierten studentischen Arbeitsgruppen.

Eine Reihe von Untersuchungen beschäftigt sich mit dem Einfluß unterschiedlicher Prüfungsformen. Die Forschungslage ist nicht ganz geklärt, doch scheinen Aufsätze (essays) die Konzentration des Arbeitsverhaltens auf wesentliche Punkte, objektive Tests das Lernen von Einzelheiten zu bewirken (BALCH 1964; nur teilweise bestätigend: HAKSTIAN 1971).

Als quantitative Wirkung auf die kognitiven

Strukturen selbst wird die Anhebung des Wissensniveaus (DRESSEL 1961) wie die Festigung von Kenntnissen (PAULSEN 1902) vermutet, während in negativ bewerteter Richtung die Prüfung als Ursache der Reduzierung der Leistung auf ein Mittelmaß erscheint (JONES 1964/65).

Vorliegende empirische Untersuchungen konzentrieren sich auf die leistungssteigernde Wirkung gehäufter Testanwendung. Von FITCH et al. (1951) wird die leistungssteigernde Wirkung häufiger Tests bestätigt; NOLL (1939) kann eine solche Wirkung nur für die leistungsschwächeren Studenten nachweisen, während HAKSTIAN (1971) keine leistungssteigernde Wirkung findet. Andere Untersuchungen belegen, daß Aufsätze (essays) objektiven Testverfahren in der Wirkung auf Leistungssteigerung und Behalten überlegen sind (MEYER 1934), was jedoch in einer für Schulverhältnisse repräsentativen Verwendung nicht bestätigt werden kann (VALLANCE 1947). Die unterschiedlichen Ergebnisse dürfen bei der verschiedenartigen situativen Einbettung der Testverfahren nicht verwundern. Die situative Distanz von Tests zu den hier anstehenden zusammengefaßten Prüfungen (comprehensive examinations) läßt die schon in sich widersprüchlichen Resultate kaum noch als Erkenntnisgewinn erscheinen.

Eine empirische Erhebung im Rahmen des einleitend erwähnten Forschungsprojekts ermittelt aus der schriftlichen Befragung von Studenten bei Vordiplomprüfungen in den Fächern Psychologie und Pädagogik, daß (im Urteil der Studenten) die Prüfung zu einer im Vergleich zum bisherigen Studium größeren Erweiterung des Wissens veranlaßt (ALVES & LODDERS 1974).

Als qualitative Wirkung auf die kognitiven Strukturen wird eine Zunahme der Integration des Stoffes (DRESSEL 1961; OPPENHEIM et al. 1966/67) vermutet; dem steht auf der negativen Seite gegenüber, daß in der Vorbereitung auf Prüfungen nur abfrag- und lernbares Wissen erworben wird (PAULSEN 1902), die Strukturbil-

dungen also im wesentlichen im Bereich des Gedächtnisses, nicht aber in dem des Verständnisses erfolgen (KURSITZA 1967). JONES (1964/65) erwähnt das Einprägen von unsinnigen Einzelfakten in der Prüfungsvorbereitung.

Empirische Untersuchungen dazu sind kaum vorhanden. Die schon erwähnte Untersuchung (ALVES & LODDERS 1974) läßt die Studenten auch die Relevanz des erworbenen Wissens einschätzen. Ein großer Anteil des für die Prüfung erworbenen Wissens wird als nicht relevant eingeschätzt.

Der Einfluß der Prüfung auf das Verhalten muß sich nicht allein in der kognitiven Strukturierung niederschlagen, sondern die Prüfung kann auch als Einübung nicht-kognitiver Verhaltensdispositionen gesehen werden. Entsprechend wird der Prüfung in einem eher negativ bewerteten Sinn die Funktion zugeschrieben, zur Verinnerlichung einer Wettbewerbshaltung beizutragen (PRAHL 1976) oder eine Anpassungshaltung zu erzeugen (parrot ability; BLACK 1968). Einflüsse auf emotionale Verhaltensstrukturen lassen sich auch aus den Untersuchungen über Prüfungsangst ableiten. Zu den erstgenannten Aspekten, die nicht einfach empirisch zu erfassen sind, fehlen bei der insgesamt ungenügenden Forschungslage verständlicherweise empirische Untersuchungen völlig.

Abschließend ist zu betonen, daß die Transformation der Sozialisations/Normierungsfunktion in die Steuerungsfunktion sichtlich eine Problemverengung darin darstellt, daß allein die direkten Einflüsse der Prüfung auf die Studenten thematisiert werden, während die soziologische Funktionsbetrachtung sich gleichfalls mit den Einflüssen der Prüfung auf die Prüfer und die Hochschule als Institution beschäftigt, die wiederum indirekt über die Struktur des Studiums usw. auf die Studenten einwirken. Für den Gang dieser Arbeit aber ist die direkte Wirkung zentraler. Der Vollzug der Prüfung baut im wesentlichen auf der Prüfungsvorbereitung auf,

die sich deutlich aus dem Studium heraushebt. Der Einfluß des Vollzugs der Prüfung auf den Prüfling ist insofern über die Prüfungsvorbereitung vermittelt. Er kann nur in dem Maße pädagogisch sinnvoll sein, in dem die Prüfungsvorbereitung als Lernprozeß und als Strukturbildung das ist. Und daran zumindest lassen die aufgezählten begrifflichen und empirischen Belege zweifeln. (In Kapitel III.I. wird anläßlich der Untersuchung der Validität der Notengebung das noch genauer erörtert.)

Die Diagnosefunktion der Prüfung

An dieser Stelle hat die Darstellung über das analytisch-berichtende Vorgehen hinaus vermehrt zielbezogene Überlegungen einzubeziehen, da die Wirkung des Vollzugs der Prüfung auf den Prüfling dieser Diagnosefunktion zuzuordnen ist. Quelle solcher zielbezogenen Forderungen an die Prüfung stellen die Konzepte pädagogischer Diagnostik dar, die in der letzten Zeit entwickelt wurden. Konkrete Zielbestimmungen allerdings sind der Systematik des Vorgehens (als Einkreisung der thematischen Funktion Rückmeldung) entsprechend erst im folgenden Teil zu finden.

Die Betrachtung der Prüfung als Diagnose steht in enger Verbindung mit der Entwicklung der psychologischen Diagnostik als Wissenschaft. Mit dem Versuch der Objektivierung der psychologischen Diagnostik ist die Einführung von methodischen Kriterien für Diagnosen verbunden. Die der Testforschung entstammenden methodischen Kriterien (Gütekriterien) wie Objektivität, Zuverlässigkeit und Validität wurden unter der strukturellen Gleichsetzung von Prüfungen und Tests auch auf Prüfungen angewendet. Eine Vielzahl von Untersuchungen in den 30er Jahren, dem ersten Höhepunkt der empirischen Prüfungsforschung (INGENKAMP 1975), stellt übereinstimmend die mangelnde Erfüllung solcher Kriterien durch

die bestehenden Prüfungen fest. Untersuchungen bis auf den heutigen Tag bestätigen das. Entsprechende Ergebnisse werden in Kapitel III.I. im einzelnen berichtet.

Als Versuch, die unkritische, entscheidende Verwendung von Prüfungsergebnissen, insbesondere für Selektionen, zu verunsichern, ist die Gleichsetzung von Prüfungen und Testverfahren (BIÄSCH 1968; KVALE 1972; SCHEER & ZENZ 1973) sinnvoll. Für eine umfassende Analyse der diagnostischen Funktionen von Prüfungen aber sind die rein methodischen Kriterien, auch bei weitem normativem Verständnis unter Einbeziehung etwa der Konstruktvalidität, unzureichend. In der psychologischen Diagnostik werden die Kriterien gleichfalls in ihrer Bedeutung eingegrenzt. Die psychologische Diagnose wird nicht mehr allein unter dem Anspruch betrachtet, eine an methodischen Kriterien ausgerichtete Datenerfassung zu sein, worauf sie die Testforschung zu reduzieren trachtet. Demgegenüber wird der sozialpsychologisch prozessuale Aspekt der diagnostischen Tätigkeit betont (HOLZKAMP 1966), der zwischen Diagnostiker und Auftraggeber stattfindet. Die Diagnostik wird in ein Handlungsschema eingebettet (KAMINSKI 1970), in dem sie an konkrete Fragestellungen und Verwendungen angebunden wird. Die amerikanische Forschung hatte schon früher verdeutlicht, daß auch die Gütekriterien in Abhängigkeit von den Verwendungszusammenhängen der Diagnosen zu sehen sind (CRONBACH & GLESER 1957).

Von der erkenntnisleitenden Bedeutung des Verwendungszusammenhangs der Diagnosen muß gleichfalls die Formulierung der Diagnosefunktion von Prüfungen ausgehen. Genau das wird von den Ansätzen der pädagogischen Diagnostik (HOPF 1975; INGENKAMP 1975) getan, wenn verschiedenartige Verwendungen von Diagnosen und die daraus erwachsenden Anforderungen an die Binnenstruktur des diagnostischen Prozesses herausgearbeitet werden. Die Anwendungen wiederum sind aus Zielen pädagogisch sinnvollen Handelns zu gewinnen, was trotz

des umfassenden Anspruchs nur allzuleicht durch die traditionelle Beschränkung auf technische Probleme verdeckt wird (wie in den Lehrbüchern des educational measurement). Trotz der Negativbeispiele solcher Lehrbücher hat die pädagogische Diagnostik in den USA eine längere Tradition (INGENKAMP 1975) und ist in der Verbindung von Diagnose und Beratung bzw. Anleitung (guidance) schon wieder einen Schritt über eine reine Diagnostik hinaus.

Die Systematik der Verwendung pädagogischer Diagnosen geht gewöhnlich von den Adressaten der Diagnosen aus; die folgende Aufzählung basiert auf SCHREINER (1970). Sie wird inhaltlich ausgefüllt durch die von HOPF (1975) dargestellten konkreten Funktionen. Dabei sind Umformulierungen notwendig, weil die beiden Arbeiten auf die Bedingungen des Schulsystems ausgerichtet sind.

An den Adressaten Student richtet sich die Informationsfunktion der pädagogischen Diagnostik, die den Lernenden über Lernergebnisse und Lernmöglichkeiten informieren soll (HOPF 1975). Von den beiden möglichen Inhalten der Mitteilung berücksichtigt die gegenwärtige Prüfung sicher nur die (abschließende) Darstellung der Lernergebnisse. Sie ist Gegenstand dieser Arbeit.

Der Lehrende ist Adressat von zwei verschiedenen Funktionen.

1. Die in der pädagogischen Diagnostik gewonnenen Daten über Merkmale des Lernenden, über Lernbedingungen und Kriteriumsleistungen sollen eine Zuordnung der Lernenden zu differenzierten Lernbedingungen ermöglichen, die an der Kriteriumsleistung überprüft werden kann. Die Zuordnung kann erfolgen durch Klassifikation oder durch Anleitung (guidance; KANDEL 1936; SIEGEL 1958). Für Beratungen und Anleitungen ist eine Sammlung von Daten über die Lernenden nützlich (DRESSEL 1961;

OPPENHEIM et al. 1966/67). Abschlußprüfungen kommen dafür prinzipiell nicht in Frage; Zwischenprüfungen dürften nur unter besonderen Umständen so verwertbar sein, zumal auch sie eher abschließende Bedeutung haben.

2. Die Prüfungen sollen dem Lehrenden Aufschluß geben über das Erreichen der gesetzten Ausbildungsziele (SADER et al. 1970). Eine Wirkung der Prüfung im Sinne dieser Funktion wäre möglich. Doch ist anzumerken, daß die Prüfer diese Sichtweise nicht teilen (TIMMERBRINK 1971) und deshalb die Verwirklichung der Funktion fraglich ist.

Für das Bildungssystem soll die pädagogische Diagnostik einen Beitrag zur Überprüfung der Curricula (DRESSEL 1961) oder zur Evaluierung des Systems als Ganzem (KVALE 1972) leisten. Auch diese Leistung scheinen die gegenwärtigen Prüfungen nicht zu erbringen. Die Einstellungen der Prüfer (s.oben) sprechen dagegen; allerdings sind Ansätze zu finden in dem Versuch der Universitäten, sich gegen den Vorwurf der Leistungsabnahme mit den Verteilungen der Prüfungsnoten zur Wehr zu setzen.

Schließlich soll die pädagogische Diagnostik für die Gesellschaft, die allerdings noch in unterschiedliche Interessensinstanzen differenziert werden müßte, eine Leistungsprognose abgeben (SCHREINER 1972) und die Erreichung von Ausbildungszielen dokumentieren (MILLIS 1966). Damit ist u.a. die Auslese- und Rekrutierungsfunktion (SCHEER & ZENZ 1973) angesprochen. Sie hat zentrale Bedeutung für die gegenwärtigen Prüfungen, wie die soziologische Analyse nachwies.

Bezogen auf die bestehenden Prüfungen kommen die beiden Übersichten von HOPF (1975) und INGENKAMP (1975) zum Schluß, daß die bestehenden Prüfungen nicht einmal für die Selektion geschweige denn für andere Funktionen

der pädagogischen Diagnostik geeignet sind. Wie angedeutet, sind bestimmte Anforderungen schon aus strukturellen Gründen, der Anordnung der Prüfung im Studium, nicht zu erfüllen. Beide Autoren betonen übereinstimmend, daß Untersuchungen über Wirkungen der Prüfungen im Sinne der funktionalen Ziele pädagogischer Diagnostik wie auch Untersuchungen über mögliche Nebenwirkungen von Prüfungen weitgehend fehlen. Es liegen in Fülle allein die in der Einleitung des Abschnitts erwähnten Qualifizierungen von Prüfungen in den methodischen Kriterien der Testforschung vor. Die sind zweifellos eine Grundlage möglicher Wirksamkeit der Prüfungen als Instrumente pädagogischer Diagnostik; die Kriterien wären aber je nach Aufgabenstellung ganz unterschiedlich zu gewichten. Hinzukommen müßten außerdem Untersuchungen über die praktische Verwendung der Prüfungsergebnisse. INGENKAMP (1975) deutet die Fülle von Tätigkeiten, die bei einer praktisch ausgerichteten Verwendung pädagogischer Diagnostik anfallen, in folgender Aufzählung an: Vergleichen, Analysieren, Prognostizieren, Interpretieren, Informieren und Beobachten von Wirkungen.

Bei allen Aufzählungen von funktionalen Aspekten sollte ein weiteres Ziel nicht übersehen werden: Die pädagogische Diagnostik hat von der Maxime auszugehen, sich auf das notwendige Mindestmaß zu beschränken, um nicht zur pädagogischen Forderung nach Selbststeuerung des Lernens in Widerspruch zu geraten. Für den tertiären Bereich ist diese Forderung besonders zentral. Die Untersuchung der Funktion der Prüfung, als Mittel pädagogischer Diagnostik die Studenten über Lernergebnisse zu informieren (im folgenden: Rückmeldefunktion), knüpft hier normativ an. Die Rückmeldefunktion ist mit der Selbststeuerung der Lernenden verbunden, indem Informationen über Lernergebnisse in die erkennende und handelnde Verfügung der Lernenden selbst übergehen.

Die Auswahl gerade dieser Themenstellung muß deshalb angesichts der allgemein unzureichenden Forschungslage bei der Diagnosefunktion wie auch bei großen Teilen der Steuerungsfunktion nicht als willkürlich erscheinen. Sie betont gegen die fremdbestimmende Auslese und Normierung durch die Prüfung einen Aspekt, bei dem Selbstbestimmung notwendiger Bestandteil ist.

Zusammenfassung (II.2.)

Eine historisch-soziologische Funktionsanalyse der Prüfungen schreibt der Prüfung als Vermittlung von Universität und Gesellschaft die beiden Funktionen der Selektion (Auslese) und der Sozialisation/Normierung zu. Beide Funktionen sind an Bedeutung eher im Zunehmen begriffen, weil der gesellschaftliche Einfluß auf die Institutionen der tertiären Bildung zunimmt. Das verstärkt die Konflikthaltigkeit von Prüfungen, die ohnehin schon vorhanden ist, weil die beiden Hauptfunktionen sich aus in unterschiedliche Richtungen weisenden Teilfunktionen zusammensetzen.

Eine Transformation auf die psychologische Betrachtungsebene verdeutlicht die Vielfältigkeit noch. Eine große Zahl von begrifflichen Analysen, der nur wenige eigenständige empirische Ansätze sich zugesellen, zeigt, daß die psychologische Transformation der Sozialisationsfunktion, die Steuerungsfunktion, als Einfluß der Prüfung auf das Lernverhalten der Studenten vor der Prüfung nur zweifelhaft erfüllt wird. Die psychologische Transformation der Auslesefunktion stellt die Diagnosefunktion dar, die in Anlehnung an die Konzepte der pädagogischen Diagnostik eine Erweiterung der Auslesefunktion mit sich bringt. Neben Lehrenden, Bildungssystem und Gesellschaft wird der Student selbst als Adressat der Diagnose bestimmt. Die Rückmeldung des Lernerfolgs an den Studenten als Thema der Arbeit wird als Beitrag zur Verwirklichung des

selbstbestimmten Lernens und Handelns der Studenten verstanden.

II.3.: Rückmeldefunktion der Prüfung

Der Begriff der Rückmeldung (Rückkopplung) entstammt den Denkmodellen der Kybernetik, wird aber unspezifisch auf einen Vorgang angewendet, in dem Informationen über das Ergebnis von Handlungen, insbesondere Lernhandlungen, an die handelnde Person herangetragen werden. Dabei werden einzelne kurzzeitig erstreckte Lernprozesse nicht von langzeiterstreckten Lernsequenzen unterschieden, obwohl die Art der Rückmeldung bei beiden unterschiedlich ist. An dieser Stelle geht es um die Rückmeldung bei Lernsequenzen, der der Begriff der summativen Evaluation (BLOOM et al. 1971) entspricht, der aber nicht nur auf Personen, sondern auch auf Institutionen bezogen wird (WULF 1972).

Die Frage, ob die Prüfung eine ausreichende Rückmeldung für den Prüfling liefert, ist nur dann zu beantworten, wenn für die Wirkung der Rückmeldung Ziele (Normwerte) formuliert werden. Dabei soll von vornherein beachtet werden, was abschätzbar in den Rahmen möglicher Wirkungen fallen kann (wobei diese "Realistik" gleichfalls normativ zu begründen ist). Ein erster Schritt ist mit der Einschränkung auf den Abschluß von Lernsequenzen getan. Weitere Einschränkungen ergeben sich für die Prüfung als Grundlage möglicher Rückmeldung aus ihrer Einbettung in das Studium und aus den Gegenständen und Inhalten der Rückmeldung.

Struktur, Gegenstand und Inhalt der Rückmeldung

Kennzeichnend für die gegenwärtigen Prüfungen ist, daß sie (im Falle des Bestehens) einen definitiven Abschluß für eine längere Phase des Studiums (oder für das ganze Studium) darstellen. Als ein solcher Abschluß sollten sie sich auf das gesamte vorhergehende Studium

beziehen (SADER 1968; Kongruenzforderung); sie geben dann Informationen über Erfolg und Mißerfolg in dem entsprechenden Studienabschnitt oder im gesamten Studium. Einschränkend ist anzumerken, daß die Prüfungen weitgehend zur Etablierung einer intensiven, selbständigen Phase der Prüfungsvorbereitung geführt haben (ALVES & LODDERS 1974; s.oben Steuerungsfunktion). Das kann sich beeinträchtigend auf die Validität der Rückmeldung auswirken. Hier wird erst einmal angenommen, daß sich die Gegenstände der Rückmeldung nicht danach unterscheiden, ob sie Lernprozessen im Studium oder in der Prüfungsvorbereitung als Teil des Studiums entstammen.

Der Bezug zu den vorhergehenden Lernprozessen, und damit der Gegenstand der Rückmeldung, wird hergestellt durch den Erwerb wissenschaftlich ausgewiesener Konzepte in einem strukturierten System (im folgenden kurz: Strukturbildungen). Als kognitive Strukturbildungen können sie formal näher gekennzeichnet werden durch solche Begriffe wie "Wissen über ", "Verständnis von" usw. Es gehören dazu aber auch affektiv-soziale Strukturbildungen, wie beispielsweise die Bereitschaft zu wissenschaftlich gerechtfertigtem Handeln. Beide Arten sind vielfach miteinander vermischt. Im Zusammenhang mit der Bewertung des Studienerfolgs haben BLOOM (1972) für den kognitiven Bereich, im gleichen Zusammenhang KRATHWOHL et al. (1964) für den affektiven Bereich Taxonomien solcher Strukturbildungen vorgelegt. Ein Blick auf die Realität der Prüfungen läßt jedoch die vorgeschlagenen Differenzierungen nicht als brauchbar erscheinen.

Prüfungen mit Abschlußcharakter (comprehensive examinations; DRESSEL (1961)) erheben eine punktuelle Leistungsstichprobe (Ausnahme: schriftliche Hausarbeiten). Die Stichprobe ist nach inhaltlichen Unterteilungen in Fachgebiete (Einzelfächer) organisiert. Die Beurteilung erfolgt im Hinblick auf die Güte der Leistung,

dem Anspruch nach also allein auf eine einzige Dimension der möglichen Qualifizierung kognitiver Strukturbildungen beschränkt. In der Form einer Note wird die Beurteilung den Prüflingen übermittelt (rückgemeldet).

Der Inhalt der Rückmeldung als Leistungsgüte von kognitiven Strukturbildungen, gegenständlich differenziert allein nach Strukturbildungen in den einzelnen durch die Prüfungsordnung vorgeschriebenen Prüfungsfächern, ist damit gegenüber möglichen Zielsystemen (Taxonomien) wissenschaftlicher Strukturbildung zweifach reduziert: auf kognitive Strukturbildungen beschränkt und auf eine einzige Dimension gegenüber einer Vielzahl von möglichen abgemagert. In dem anschließenden Teil zum Leistungsbegriff soll verdeutlicht werden, unter welchen Umständen eine solche Reduktion gerechtfertigt erscheint.

An dieser Stelle ist zur Herausarbeitung der Themenstellung weiter zu fragen, welche möglichen und sinnvollen Wirkungen die Rückmeldung als Beurteilung der Leistungsgüte haben kann. Die Frage richtet sich an die Zeit nach der Prüfung. Dabei unterscheiden sich Zwischenprüfungen (Vordiplom) von Endprüfungen (Hauptdiplom); in einer rein formalen Betrachtung jedoch verwischen sich die Unterschiede.

Nach der Prüfung ist zum einen die Fortführung von auf Erwerb kognitiver Strukturbildungen im Bereich wissenschaftlichen Wissens gerichteten Lernsequenzen möglich. Dies kann geschehen durch den Beginn neuartiger Lernsequenzen, die an die Lernsequenzen vor der Prüfung mehr oder weniger eng anknüpfen, oder in Form der Wiederholung von Lernsequenzen, die die Strukturbildungen vor der Prüfung nur modifizieren. Beispiel für das letztere ist die Wiederholung von Kursen, nachdem sich in Prüfungen herausgestellt hat, daß die Strukturbildungen bestimmten Anforderungen nicht entsprechen. Eine neuartige Strukturbildung ist die Aufnahme eines Aufbaustudiums nach der Endprüfung. Zum

anderen ist das Nachher gekennzeichnet durch die Anwendung von Strukturbildungen im Sinne der Aufnahme einer praktischen Tätigkeit, die entsprechende kognitive Strukturbildungen voraussetzt. (Es sollte deutlich geworden sein, daß Wissen und kognitive Strukturbildung sehr weit gefaßt werden müssen und, um für eine wissenschaftliche Ausbildung repräsentativ zu sein, auch z.B. Wissen über Anwendung umfassen sollten. Auch darüber müßten Urteile in Form der Leistungsgüte möglich sein. Ob aber die verlangten Strukturbildungen überhaupt in der Leistungsstichprobe vertreten sind, ist eine andere Frage. Sie wird in Kapitel III.I. erörtert.)

Das Nachher der Prüfung beinhaltet also vereinfacht eine planende Aktivität des Studenten in bezug auf folgende Strukturbildungen (Lernprozesse) oder Anwendungen der erworbenen Strukturbildungen (praktische Tätigkeiten). Die planende Aktivität sollte sich als oberstes Ziel die gesellschaftlich wie individuell rationale Passung von individuellen Leistungsmöglichkeiten und sozialen Anforderungen setzen. Der Begriff der Passung ist dabei komplex zu sehen. Er könnte mit dem PIAGETschen Begriff der Anpassung verglichen werden, die als Einheit von Akkommodation und Assimilation eine wechselseitige Bezogenheit meint. So fordert Passung von individuellen Leistungsmöglichkeiten und gesellschaftlichen Anforderungen eben nicht allein die schiere Anpassung von Individuen an vorgegebene gesellschaftliche Voraussetzungen, sondern gleichberechtigt die Möglichkeit der Änderung der Anforderungen im Sinne der Leistungsmöglichkeiten als individuellen Voraussetzungen. Ferner fällt darunter die geplante Änderung von eigenen Leistungsmöglichkeiten im Hinblick auf bestimmte Anforderungen.

Ein solcher Begriff der Passung hat zentrale Bedeutung auch für die pädagogische Diagnostik (KANDEL 1936; HOPF 1975).

Der Beitrag der Rückmeldung als Beurteilung der

Leistungsgüte der vor der Prüfung erworbenen kognitiven Strukturbildungen richtet sich sowohl an die kognitiven als auch an die dynamischen Grundlagen einer über Handlungsplanungen und Selbsterkenntnis zugleich vermittelten Passung. Als die entsprechenden Prozesse sollen Selbstbewertung der Studenten und Motivation der Studenten betrachtet werden.

Die Selbstbewertung meint die Bewertung des Zustands der individuellen Strukturbildungen (Leistungen) durch den Prüfling. Sie stellt die mehr kognitiv-rationale Komponente der Passung dar. Ein Einfluß der Prüfung auf die Selbstbewertung wird von vielen Autoren positiv wertend vermutet oder angestrebt (AGER & WELTMAN 1966/67; HATCH 1958; HOPF 1975; KURSITZA 1967; KVALE 1972; SADER et al. 1970; WILHELMS 1967). Auch die geforderte oder vorausgesetzte Selbstkontrolle des Lernerfolgs durch Prüfungen kann in dieser Richtung interpretiert werden (EBEL 1958; HIMMELWEIT 1966/67; KAPUSTE 1970; MILLER 1964; SALZMANN 1966; SCHREINER 1970).

Die Motivation meint die Bereitschaft zur Aktivität in bezug auf neuartige oder Verbesserung alter Strukturbildungen in Form von Lernsequenzen oder die Anwendung der erworbenen Strukturbildungen in praktischen Tätigkeiten. Sie stellt die mehr volitiv-aktive Komponente dar (wobei das "mehr" andeuten soll, daß es sich hier möglicherweise um eine zwar übliche, aber problematische Trennung handelt, die keinesfalls als die von kognitiven und emotionalen Prozessen gesehen werden darf. Handlungsplanungen beinhalten kognitive Prozesse, und Selbstbewertungen sind nicht emotionsfrei.)

Normwerte der Selbstbewertung

Das Konzept der Selbstbewertung ist eng mit der

Wertvorstellung von der Autonomie des lernenden Individuums verbunden. Der Beitrag der Rückmeldung zur Selbstbewertung soll als Aufklärung über sich selbst das Maß der individuellen Freiheit vergrößern (WILHELMS 1967). Die Freiheit besteht darin, die Passung auf der Grundlage einer realistischen Selbstbewertung in unterschiedlicher Richtung vorzunehmen: Sie liegt bis zu einem bestimmten Grad in der Verfügung des Individuums.

Als Normwert für den sinnvollen Einfluß der Prüfung auf die Selbstbewertung ist entsprechend die Realistik der Selbstbewertung zu fordern. Realistik heißt, daß die Selbstbewertung den wirklichen Leistungsmöglichkeiten entspricht. Nur dann kann die Passung erfolgreich sein. (Die Passung selbst kann, wenn sie aktiv als erhöhter Einsatz zu erhöhten Leistungen führt, wieder eine Änderung der Selbstbewertung bewirken.)

Der Normwert ist über die Richtung des Einflusses (mehr Realistik) hinausgehend auch für das Ausmaß des Einflusses zu formulieren. Da die Kompetenz in den Fächern der Psychologie und in der Psychologie insgesamt sicher nicht durch die Leistungsstichprobe in der Prüfung voll abgedeckt wird, wäre eine vollständige Kongruenz von Leistungsbewertung und Selbstbewertung nach der Prüfung nicht angemessen. Es gehört mit zur Freiheit des lernenden Individuums, der Rückmeldung in Prüfungen einen begrenzten Stellenwert zuzuschreiben. (Ergänzend muß auch noch eine Kontrolle der Qualität der Rückmeldung selbst erfolgen, die hier vorausgesetzt wird.)

Eine Quantifizierung des als Normwert zu erwartenden Ausmaßes der Wirkung überfordert die theoretischen Vorüberlegungen. Statt dessen soll eine weitere Qualifizierung versucht werden: Die dynamische Interpretation der Passung, aber auch schon ein nicht statisches Konzept von kognitiven Strukturbildungen lassen eine Einschränkung der Wirkung auf die Selbstbewertung des gegenwärtigen Zustands der Strukturbildungen sinn-

voll erscheinen.

Zusammenfassend ist also zu erwarten: Die Prüfung beeinflußt bedeutsam die Selbstbewertung in Übereinstimmung mit den Ergebnissen. Eine vollständige Kongruenz von Ergebnis und Selbstbewertung ist allerdings nicht als sinnvoll anzusehen. Die Wirkung auf die Selbstbewertung sollte sich mehr auf den gegenwärtigen Zustand beschränken.

Normwerte der Motivation

Im Hinblick auf die Passung von Individuum und gesellschaftlichen Anforderungen verschiebt sich das Gewicht bei der Motivation mehr in Richtung der gesellschaftlichen Anforderungen. Eröffnet die realistische Selbstbewertung neue Freiheiten in reflektierter Form, so ist eine Vergrößerung der Autonomie über die Motivation nur indirekt möglich. Die Umsetzung der Rückmeldung in Handlungsplanung kann rein funktional erfolgen, wenngleich reflektierte Zwischenschritte bei zeitlich weitgreifender Planung wohl zunehmend erforderlich sind (doch sie sind wiederum im Hinblick auf das Planungsziel funktional). Normwerte für die Motivation lehnen sich entsprechend enger an gesellschaftliche Anforderungen und sie vermittelnde institutionelle Zwecke an.

In Abweichung von der Selbstbewertung ist bei der Motivation eine Normierung in zwei Richtungen möglich. Auf der einen Seite sollte die Handlungsplanung der Rückmeldung in einem positiven Sinn entsprechen. Die Motivation sollte mit der positiven Ausprägung der Rückmeldung zunehmen. Diese Forderung ist von den gesellschaftlichen Anforderungen her sinnvoll, weil die individuellen Fähigkeiten dadurch am besten genutzt werden können. Zum anderen sollte die Motivation mit der negativen Ausprägung zunehmen, indem eine schlechte Leistung in der Prüfung durch einen zusätzlichen

Einsatz kompensiert wird. Das Nichtbestehen sei ausgeklammert. Es hat gemäß der Prüfungsordnung eine Zwangskompensation zur Folge, die Wiederholung der Prüfung.

Beides zusammen ergibt als Normwert eher eine kurvilineare Beziehung. Auch hier gilt wieder die Einschränkung, daß die Beziehung nicht eine vollkommene Abbildung sein sollte, u.a. weil sie in weitere Rückmeldungen eingebettet ist. Im Vergleich zur Selbstbewertung sollte die Enge der Beziehung bei der Motivation eher geringer sein, weil hier noch zusätzlich andere Determinanten individueller Handlungsplanung wie sachlich orientierte Interessen genügend Raum haben sollten.

Die unterschiedliche Form der normativ geforderten Beziehung von Rückmeldung zu Motivation und Selbstbewertung läßt keine bedeutsame Zunahme der Abhängigkeit zwischen den beiden abhängigen Prozessen erwarten. Eine normative Abwägung des Zueinanders der beiden abhängigen Variablen beschränkt sich nicht auf deren empirische Relation, sondern hat auch eine Bedeutungsgewichtung vorzunehmen. Dieser zufolge ist die Selbstbewertung eindeutig mit mehr Gewicht zu versehen. Nur sie berücksichtigt zweifelsfrei die Forderung der pädagogischen Diagnostik, daß die Rückmeldung einen Beitrag zur Selbststeuerung des Lernens zu leisten habe. Die Selbststeuerung setzt eine rationale Selbsteinsicht voraus, die die geforderte Verarbeitung der Rückmeldung in der Motivation nicht notwendig umfaßt. Die Vermittlung der Rückmeldung über die Selbstbewertung stellt ein Gegengewicht dar gegen die zunehmende Betonung fremdbestimmender Ausleseprozesse in Prüfungen, die nicht auf Selbststeuerung vertrauen. Dem folgenden Teil bleibt in diesem Zusammenhang ein Eingehen auf die von KVALE (1972) aufgeworfene Frage überlassen, ob die beabsichtigte Vermittlung über die Selbstbewertung nicht nur eine Verschleierung der Herrschafts-

funktion von Prüfungen durch Humanisierung darstellt.

Der Versuch einer Bestimmung von Normwerten hat nur allgemeine Vorgaben erarbeiten können, die zur Diskussion auffordern. Die ermittelten Normwerte sind nicht genügend konkret (und quantifiziert), um direkt in der Auswertung verwendet zu werden. Sie sind auch nicht erschöpfend genug, um als einzige Leitlinie der begrifflichen und empirischen Analyse zu dienen. Das scheitert u.a. daran, daß sie selbst wiederum von den Ergebnissen der begrifflichen und empirischen Analyse nicht unabhängig sind: Beispielsweise müßte das mögliche Verhältnis von Selbstbewertung und Motivation analytisch differenzierter untersucht werden, bevor Normwerte darüber aufgestellt werden.

Zusammenfassung (II.3.)

Die Rückmeldung, die die Prüfung wird leisten können, wird eingegrenzt auf die Mitteilung von Erfolg und Mißerfolg einer Lernsequenz. Inhalt ist eine Mitteilung über die Leistungsgüte der kognitiven Strukturbildungen, differenziert nur nach verschiedenen Prüfungsfächern. Die Ziele der Rückmeldung gehen zurück auf die Passung von individuellen Leistungsmöglichkeiten und objektiven Anforderungen. Diese Passung ist vermittelt über die Selbstbewertung und Motivation der Prüflinge nach der Prüfung.

Als Normwert für die Selbstbewertung ist deren Realistik zu fordern, die sich in einem bedeutsamen Einfluß der Prüfung in Richtung einer Übereinstimmung von Rückmeldung und Selbstbewertung ausdrücken sollte. Der Einfluß sollte eher auf die gegenwärtige Selbstbewertung beschränkt sein.

Als Normwert für die Motivation (Handlungsplanung) wird eine kurvilineare Beziehung gefordert, weil gesellschaftlich sinnvoll sowohl die besonders gute

Leistung zu erhöhter Aktivität als auch die schlechte Leistung zu kompensatorischer Aktivität führen sollten.

Die Abwägung von Selbstbewertung und Motivation gegeneinander läßt die Selbstbewertung als die wesentlichere abhängige Variable erscheinen, weil durch sie Fremdbestimmung in Richtung auf Selbstbestimmung begrenzt werden kann.

II.4.: Leistungsbewertung und Leistung

Dieser Teil verfolgt zwei Ziele. In einem ersten Schritt soll ein Thema nachgeholt werden, das bei der Beschäftigung mit Prüfungen kaum zu übersehen ist, im Hinblick auf die Rückmeldefunktion zudem zentral ist: Noten als institutionell vorgegebenes Mittel der Rückmeldung des Prüfungserfolgs an die Prüflinge.

In einem zweiten Schritt sind die Implikationen, die sich daraus ergeben, daß die Rückmeldung über die Noten ein problematisches "Leistungsprinzip" einführt, zu diskutieren. Dazu trägt ein Anknüpfen an die Auseinandersetzung um das Leistungsprinzip und die Notwendigkeit von individuellen vergleichsorientierten Leistungsnachweisen bei.

Leistungsbewertung durch Noten

Noten sind mit den gegenwärtigen Prüfungen sowohl in der Schule wie besonders in der Hochschule unlösbar verbunden. In vielen Texten über Prüfungen könnte man den Begriff der Prüfung durch den der Notengebung ersetzen. (Umgekehrt ist das nicht immer der Fall, weil in der Schule Noten auch ohne Prüfungen vergeben werden.)

Noten vermitteln die wesentlichen zukünftigen Folgen der Prüfungen (Diagnosefunktion); die Steuerungswirkung wird zwar nur indirekt, aber gleichfalls durch sie beeinflußt. Dementsprechend sind die Noten wie die

Prüfungen Funktionenmischungen. Das geht aus der Aufzählung von Notenfunktionen in den gängigen Werken zum educational measurement hervor, die sich wegen ihres meßtechnischen Vorbehalts weniger mit Prüfungen beschäftigen, sondern mit der Verwandlung von Testergebnissen in Noten (BROWN 1971; NUNNALLY 1972; THORNDIKE & HAGEN 1969). Die Polifunktionalität wird selbst dort noch vertreten, wo Noten durch Lerndiagnosen ersetzt werden sollen (SCHREINER 1970). Das übliche Nebeneinander der Aufzählung von Notenfunktionen ist unbefriedigend. Mehr Klarheit über das Verhältnis der Funktionen zueinander ist aus der Geschichte der Noten zu gewinnen. Die bestätigt das, was aus der engen Bindung von Prüfung und Noten zu erwarten ist: die Dominanz der Auslesefunktion auch bei den Noten.

In der schulischen Verwendung der Noten steht nach DOHSE (1967) am Anfang der Notengebung eine Mischung von Anreiz (pädagogischer Funktion) und Auslesefunktion. Noten als Anreize werden in der auf Jesuitenschulen üblichen Lokation deutlich. Die Auslesefunktion der Noten tritt in den Vordergrund mit der Durchsetzung der Zonenstufung der Noten. Diese kann auf das 18. bis 19. Jahrhundert datiert werden (ZIELINSKI 1961); damit entspricht sie der bürgerlich-demokratischen Durchsetzung des Leistungsprinzips. So nimmt in der heutigen Verwendung der Noten die selektive Funktion die erste Stelle ein (DANN & MÜLLER-FOHRBRODT 1972; DOHSE 1967; DUMKE 1974); sekundär dagegen ist die pädagogische Funktion geworden. Das bestätigt der staatliche Einfluß auf das Notensystem und dessen wechselvolle Geschichte in Abhängigkeit von der jeweiligen politischen Gewalt (DOHSE 1967). Auch die Umwandlung des Prozentsystems in amerikanischen Schulen hin zum Buchstabensystem (BIXLER 1936), das sich leicht in Ziffern umformen läßt, entspricht der Kommodität selektiven Gebrauchs.

Klarer noch ist eine Einordnung der Notenverwendung in der Hochschule des kontinentalen Europa. Dort werden die Noten weniger als in der Schule unmittelbar im Unterricht benutzt und sind entsprechend direkter auf die Auslese bezogen. Nach NISBET (in: LAUWERYS & SCANLON 1969) sind Noten in akademischen Prüfungen zwar vereinzelt schon im 15. oder 16. Jahrhundert festzustellen, ihre universelle Einführung aber ist gebunden an den Übergang von der Aristokratie zur Meritokratie. In England nimmt die Einführung im Fach Mathematik (!) ihren Ausgangspunkt. Die leistungsbezogenen Konkurrenzprüfungen für den Zugang zum Öffentlichen Dienst im 19. Jahrhundert in England und die bürokratische Zentralisierung in Kontinentaleuropa fördern ihre Durchsetzung.

Gerade die Ziffernnote kommt bürokratischen Ausleseprozessen weit entgegen. KVALE (1972) bezeichnet die (Ziffern-)Note deshalb als bürokratischen Geniestreich. (Auch Buchstabennoten, wie die in den USA gebräuchlichen, werden zum Zwecke der Verrechnung als Mittelnote - grade point average - in Ziffern umgewandelt.) Die enge Anbindung an die Auslesefunktion läßt kaum erwarten, daß Noten gleichfalls auf die Anforderungen der Rückmeldefunktion passend zugeschnitten sind. Doch soll das Urteil erst auf Grund einer differenzierten Analyse gefällt werden. Für die weitere Konkretisierung des Vorgehens dieser Analyse ist festzulegen, welche Informationen prinzipiell von den Noten gegeben werden können. Einmal ist eine solche Erwartung schon eingegangen in die Bestimmung der abhängigen Variablen Selbstbewertung und Motivation (Handlungsplanung). Dann wurde, auch mit Blick auf die Noten, der Inhalt der Rückmeldung eingegrenzt auf die Mitteilung der Güte kognitiver Strukturbildungen. Darüber, daß das Ziel der Noten sein sollte, herrscht in den technisch orientierten Texten pädagogischer Dia-

gnostik weitgehend Übereinstimmung (BROWN 1971; DAVIS 1964; STANLEY & HOPKINS 1972). Es wird hier als Ziel in dem Sinne betrachtet, daß mehr von den Noten nicht erwartet werden kann.

Die Güte kognitiver Strukturbildungen (Leistungen) wird mit den Noten durch Einordnung der Leistung in eine Klasse der Leistungsgüte vorgenommen. Die verbalen Bezeichnungen der Klassen, in Prüfungsordnungen weiter ausgeführt, geben teilweise vor, absolute Bewertungen zu sein. Im Fach Psychologie an der Universität Münster enthält nur die alte Form der Prüfungsordnung Erläuterungen der Stufen. Sie sind mit Ausnahme der Note 3 ("durchschnittlich") der Versuch absoluter Definitionen. (Es ist interessant festzustellen, daß die (alte) Ordnung für Staatsprüfungen der Lehrämter rein relative Definitionen in der Prüfungsordnung vorsah.) Eine absolute Bewertung mag in den Augen der Bewerter auch angestrebt werden. Solange aber die Bewertung nicht ausgerichtet an inhaltlichen Konkretisierungen der Erwartungen erfolgt, die dem Prüfling bekannt sind (BADER 1913; KVALE 1972), hat die Bewertung der Leistungsgüte für den Prüfling eher eine relative Bedeutung durch den Vergleich mit den Bewertungen der anderen Prüflinge. Die Einschränkung zu einer relativen Information gilt um so mehr dort, wo nur eine Leistungsstichprobe Grundlage der Bewertung ist. Der relativen Bedeutung der Information entspricht der Konkurrenzcharakter (competitive character; JARRETT 1963), der ein direkter Abkömmling der Auslesefunktion ist. (Die englischen Beamtenprüfungen des 19. Jahrhunderts nahmen die Bewertung anfänglich durch Bildung einer Rangreihe vor.)

Es scheint deshalb realistisch zu sein, davon auszugehen, daß die Noten für die Prüflinge nur dadurch Bedeutung gewinnen, daß sie auf die empirische Verteilung bezogen werden. Anders als durch einen solchen Vergleich ist z.B. die Übertragung der Rückmeldung auf

künftige Anforderungen über die Selbstbewertung vermittelt nicht denkbar. Eine direkte Übertragung durch absolute Interpretation wäre nur möglich, wenn die Anforderungen für Prüfungen und für künftige Leistungen voll inhaltlich festgelegt und auf dieser Inhaltsebene vergleichbar wären (das wird im folgenden Abschnitt noch einmal aufgenommen).

Diese erneute Einschränkung auf die eher relative Bewertung der Leistungsgüte stellt die Frage, ob ein solcher Inhalt der Rückmeldung überhaupt noch vertretbar ist. Bevor das im Hinblick auf den Leistungsbegriff geschieht, soll kurz auf mögliche Entwicklungen der Notengebung eingegangen werden.

In der Erziehungswissenschaft werden immer wieder Diskussionen angestellt, die sich mit neuen Ansätzen zu Bewertungen beschäftigen und damit verbunden Änderungen in den Bewertungssystemen fordern (müssen). So schlagen erziehungswissenschaftliche Rückbesinnungen auf die pädagogische Bedeutung der Noten (FLITNER 1966; KAMPMÜLLER 1961; KARSDORF 1961; SCHREINER 1970) oder der die Bewertung als umfassende Evaluation verstehende reformerische Ansatz der amerikanischen Erziehungswissenschaft (TYLER 1942) funktional zentrierte, von der Auslese abgelöste Verwendungen von Noten, neue Methoden der Bewertung oder auch nur verbale Ergänzungen der Noten vor.

Die Rückbesinnung auf erziehungsphilosophische Grundlagen der Bewertung findet sich gleichfalls in neueren vorwiegend technisch orientierten Werken zum Messen und Bewerten wie z.B. bei THORNDIKE (1972). SMITH & DOBBIN (1960) betonen in ihrer Übersicht der Forschungslage zu Noten die zunehmende Beschäftigung mit der Beziehung von Noten und Erziehungszielen.

Doch sind auf dieser Grundlage keine umwälzenden Änderungen erfolgt oder absehbar. Zwar sind in den Grundschulen in verschiedenen Ländern die Noten abge-

schafft, aber eine Ausweitung auf die höheren Systeme ist nicht zu erwarten. So ist der Evaluationsansatz in den USA schon seit über 30 Jahren im Bereich der Notengebung ohne erkennbare Folgen. Nicht einmal in der Diskussion zeichnet sich eine klare Richtung der zielbezogenen Entwicklung der Notensysteme ab (ROSS & STANLEY 1954; SMITH & DOBBIN 1960; THORNDIKE & HAGEN 1969).

Im Hochschulbereich sind einige mehr oder weniger erfolgreiche Experimente mit P-F(pass-fail)-Systemen angestellt worden (JOHNSON & ABRAHAMSON 1968; KARLINS et al. 1969; STALLINGS et al. 1969). Das nicht so zentralistische amerikanische Hochschulsystem erlaubt sogar - schon seit längerer Zeit - ein Beispiel der Abschaffung von Noten (MARSHALL 1952; 1968). Das ändert aber nichts daran, daß die Noten die übliche Form der Bewertung bleiben. Ihre Reform ist nicht ein reines Problem einer noch so aufgeklärten Erziehungstechnologie. Die feste Verankerung der Noten in der Erziehungskultur (THORNDIKE 1972) und die emotionale Bedeutung der Noten für Lehrer und Studenten (WILHELMS 1967) sind in Rechnung zu stellen. Und es ist zu bedenken, daß die Noten genau jener Funktion von Prüfungen entsprechen, die als dominante und eher an Bedeutung zunehmende identifiziert wurde: der Auslesefunktion.

Leistungsbewertung und Leistungsprinzip

Der vorangegangene Abschnitt sollte deutlich werden lassen, daß das institutionell für die Rückmeldung vorgesehene Mittel, die Noten, nicht mehr als die vergleichende, global wertende, quantitative Einstufung der Güte einer Leistung hergibt. Es ist zu fragen, ob das von der Arbeit implizierte Selbsterleben solcher Leistung und die Ausrichtung der Handlungsplanung daran ein Ziel tertiärer Bildungsprozesse sein dürfen.

Bei der Beantwortung wird also folgender Begriff von Leistung zugrunde gelegt: Leistung als Grundlage eindimensionaler Bewertung der kognitiven Strukturierungen nach deren Güte; die Bewertung wird eher relativ erlebt im Vergleich zu Bewertungen anderer; Träger der Leistung ist immer der einzelne Prüfling; es wird nur das Produkt, nicht dessen Entstehung bewertet. Gegenüber der so verstandenen Leistung lassen sich zwei Standpunkte der Kritik anbringen, die enger, als das in Kürze angedeutet werden kann, miteinander verknüpft sind. Die vorgenommene Trennung dagegen macht die Zuordnung zur Funktionentheorie der Prüfung möglich.

Aus einer mehr psychologischen Perspektive wird im Hinblick auf die Steuerungsfunktion der Prüfung die Ausklammerung von Sinnfragen kritisiert (MITSCHERLICH 1972). Das Ziel möglichst hoher Leistungen ohne Sinnbegründung - und die ist in der Bestimmung nicht vorgesehen - erzeugt Leistungshemmungen. Die durch die Relativität der Bewertung begünstigte konkurrenzhafte (competitive) Auffassung von Leistung kann darüber hinaus eine ausgesprochene Anti-Leistungs-Haltung bedingen (COLEMAN 1959; EIGLER 1975).

Aus der soziologischen Perspektive, die sich mehr auf die Auslesefunktion richtet, aber auch auf die Steuerungsfunktion zurückblenden läßt, wird das zugrunde liegende Leistungsprinzip generell kritisiert. Das Prinzip der Zuteilung des Sozialstatus durch Leistung ist im Bereich beruflicher Tätigkeit fragwürdig, weil unmöglich geworden (OFFE 1970). Gegenüber der emanzipatorischen Bedeutung des Prinzips im Entstehen der bürgerlichen Gesellschaft hat es derzeit nur noch Verschleierungsfunktion und dient der Aufrechterhaltung der Trennung von Arbeit und Sinnlichkeit (DREITZEL 1974). Eine direkte Verbindung der Kritik zum akademischen Be-

reich entsteht dadurch, daß die erkannte Fragwürdigkeit der Leistungskontrolle im Berufsbereich zu einer Verlagerung der leistungsbestimmten Statuszuweisung in den Ausbildungsbereich geführt hat (OFFE 1970). Die in der letzten Zeit noch zunehmende Tendenz (EIGLER 1975), der die bereits beschriebene Bedeutungsvermehrung der Auslesefunktion von Prüfungen entspricht, greift verzerrend in die Bildungsprozesse ein (KLAFKI 1974), indem ein emanzipatorischer Leistungsbegriff, der den unmittelbaren Problembezug von Leistung, unter Einschluß von Gruppenleistung, hervorkehrt, unterdrückt wird.

Beide Kritiken wenden sich gegen die in Noten im Rahmen einer Prüfung erfaßbare Leistung, die vergleichsorientierte, globale Kennzeichnung der Leistungsgüte. Die Antwort darauf kann nicht erfolgen im Sinne einer naiven anthropologischen Legitimation von Leistung. Im Anschluß an KLAFKI (1974) wird folgende Antwort vorgeschlagen: Die Gesellschaft kann nicht auf die Zuordnung von Positionen und Personen durch vergleichsorientierte Leistungsbewertung verzichten, da dieses Prinzip rationaler ist als mögliche Alternativen. Die Passung von zukünftigen Anforderungen und Personen wird dort, wo die Anforderungen und Voraussetzungen nicht im einzelnen inhaltlich abgestimmt werden können, immer noch am besten über Prozesse sozialen Vergleichs gewährleistet. (Dabei muß dieser Vergleich nicht die Form fremdbestimmender Konkurrenzauslese annehmen, sondern sollte so weit wie möglich der relativen Selbstbestimmung der Individuen überantwortet werden. Erst dann wird auch eine dynamische Erfüllung dieser Aufgabe möglich, die nicht nur Auslese, sondern auch Förderung im Hinblick auf die Erreichung von Passung bedeutet.) Die Übernahme dieser Aufgabe durch die Universität sichert ihr zugleich mittels der inhaltlichen Ausgestaltung der dazu notwendigen Bewertung ein bestimmtes Ausmaß an

Einfluß auf die Gesellschaft (HONDRICH 1972). Es sollte auch nicht übersehen werden, daß für die Individuen die Einordnung in solche Leistungsnormen ein Beitrag zur Herausbildung ihrer Selbstidentität darstellen kann (BAHRDT 1972). Mit dieser Begründung aber ist eine Reihe von Forderungen über Ausgestaltung und Begrenzung dieses "Leistungsprinzips" zu verbinden, die erst den Bedenken voll Rechnung tragen.

1. Die Leistung ist so weit wie möglich mit objektiven und subjektiven Bedürfnissen als Form einbettender Sinngebung zu verbinden (EIGLER 1975; FEND 1971). Subjektiv kann die Sinngebung u.a. geschehen durch die Einbettung der Leistung in individuelle Interessenstrukturen. Eine objektive Sinngebung, die zugleich Grundlage der subjektiven darstellt, ist durch die Ausrichtung der Leistung an der Struktur der Wissenschaften zu gewährleisten. Hinzu kommt die immer wieder übersehene Einbeziehung des Praxisbezuges (BAHRDT 1972).

Leistungen im Rahmen gegenwärtiger Prüfungen tragen diesen Forderungen nur bedingt Rechnung: Eine Einbettung in subjektive Interessen ist durch beschränkte Wahlmöglichkeit der Nebenfächer wie durch die Möglichkeit der Spezialisierung innerhalb der Prüfungsfächer gegeben, doch auch prinzipiell im Hinblick auf die Vergleichbarkeit der Leistungen begrenzt. Die Orientierung an der Struktur der Wissenschaften scheint durch die Aufteilung in einzelne Prüfungsfächer erfüllt, jedoch ist nicht sicherzustellen, ob nun die Struktur der Wissenschaft die Auswahl der Fächer, oder aber die Prüfungsordnung die fachliche Gliederung der Ausbildung bestimmt. Der Praxisbezug ist zumindest im Urteil der Studenten nur unzureichend verwirklicht (ALVES & LODDERS 1974).

2. Die vergleichende Feststellung der Leistung

darf nicht Grundlage der globalen Zuweisung von gesellschaftlichem Status sein; die Statuszuweisung ist funktional zu begrenzen auf die Zuweisung von Sachautorität. Auch die Zuweisung von Sachautorität sollte in einem möglichst dynamischen Leistungsverständnis erfolgen: Sie sollte nicht rein abschließend sein, sondern Kompensationen zulassen. Die Leistungsfeststellung, die der Zuweisung von Sachautorität zugrunde liegt, sollte, um konkret interindividuelles Konkurrenzverhalten auszuschalten, in einem möglichst anonymen Leistungsvergleich erfolgen.

Die drei Forderungen bezeichnen mehr oder weniger ein über die Universität hinaus nur gesellschaftlich zu lösendes Problem. Es kann kein Zweifel sein, daß sie nur unzureichend erfüllt werden. Am weitestgehenden wird noch die dritte Forderung erfüllt, auf die die Universität auch den deutlichsten Einfluß hat.

3. Es ist ferner zu fordern, daß das "Leistungsprinzip" als Grundlage der globalen, vergleichenden Leistungsbewertung aus pädagogischen Gründen auf einen eingegrenzten Bereich im akademischen Studium beschränkt wird, um solchen Prozessen genügend Raum zu lassen, wie sie Strukturbildungen kognitiver Art, die durch mehr als durch vergleichbare Güte zu kennzeichnen sind, und Strukturbildungen affektiver, insbesondere auch sozialer Art (Gruppenleistungen) darstellen (KLAFKI 1974).

Darüber hinaus muß bei jeder vergleichenden Leistungsbewertung zur Passung von gesellschaftlichen Anforderungen und individuellen Leistungen gefragt werden, ob nicht doch eine inhaltliche, zielgerichtete Formulierung der Anforderungen und eine dementsprechende zielbezogene Leistungsmessung möglich ist. Die methodischen Vorarbeiten zur lernzielbezogenen

Leistungsmessung können dabei hilfreich sein (STRITTMATTER 1973), wenngleich ihre testpsychologische Präokkupation ein Hindernis ist.

Zusammenfassung (II.4.)

Noten und Prüfungen sind untrennbar miteinander verbunden. Die Noten sind das institutionell vorgesehene Mittel der Rückmeldung. Aus der Geschichte der Noten aber ist abzulesen, daß die Noten, insbesondere die Ziffernnoten, eng an die Auslesefunktion angebunden sind. Im Sinne der Rückmeldung geben sie eine Information über die quantitative Güte des Zustands der kognitiven Strukturbildungen (Leistungen). Die Information ist eindimensional und von den Prüflingen her gesehen eher relativ zur Verteilung der Noten bedeutsam.

Gegenüber einer Betonung konkurrenzbezogener kognitiver Einzelleistungen wird Kritik geäußert an dem Fehlen von sinnhaften Bezügen und am antiemanzipatorischen Charakter des Leistungsprinzips. Dem wird ein begrenzter Stellenwert vergleichender Leistungsbewertung bei der Passung von individuellen Fähigkeiten und gesellschaftlichen Anforderungen entgegengehalten. Die Begrenzung ist allerdings in mehrfacher Hinsicht zu fordern: als möglichst weitgehende Einbettung solcher Leistungen in objektive und subjektive Sinnzusammenhänge; als funktionale Begrenzung der Statuszuweisung auf Grund derartiger Leistungen; als Offenlassen eines weiten Bereichs von andersartigen Leistungen (auch sozialer Art), die sich einer vergleichenden Bewertung individueller Leistungsgüte entziehen. Von den gegenwärtigen Prüfungen werden diese Forderungen nur unzureichend erfüllt.

III.: Psychologische Analyse der Rückmeldefunktion

Übersicht: In diesem Kapitel sollen die begrifflichen (theoretischen) und methodischen Mittel der Psychologie eingesetzt werden, um auf dem Hintergrund der Normwerte (aber nicht vollständig an sie gebunden) die reale Wirkungsweise von Noten und Prüfungen auf Selbstbewertung und Motivation der Prüflinge erkennbar zu machen. Das soll in drei Schritten geschehen, die eigenständige Kapitel darstellen und nur locker aufeinander aufbauen.

III.I. Voraussetzungsanalyse
III.II. Wirkungsanalyse
III.III. Stufenanalyse

Die Voraussetzungsanalyse (III.I.) geht davon aus, daß die angesprochenen Variablen bisher zwar auf der Zielebene differenziert und in mögliche Verursachungszusammenhänge eingebettet worden sind, es aber an einer genaueren psychologischen Analyse, die den Status der unabhängigen und abhängigen Variablen in den Rahmen theoretischer Strukturen einordnet und in empirischen Bestimmungen konkretisiert, mangelt. In den ersten beiden Teilen des Kapitels wird das Prüfungsfeld genauer beschrieben. Die Anbindung der Wirkung an die Noten ist Anlaß für die ausführliche Beschäftigung mit den Eigenschaften der Noten in akademischen Prüfungen auf der Grundlage vorliegender Theorien und Untersuchungen. Hinzu tritt die empirische Analyse der Noten, die im Rahmen der eigenen Untersuchung anfallen. In den weiteren drei Teilen des Kapitels wird für die Variablen Selbstbewertung und Motivation die theoretische, begriffliche und empirische Tradition aufgearbeitet. Daraus wird die Begriffsbildung festgelegt und eine Operationalisierung der empirischen Untersuchung begründet, deren Ergebnisse, soweit sie sich zur Kontrolle der Erfassung des Inten-

dierten eignen, dargestellt werden.

Die Wirkungsanalyse (III.II.) beschäftigt die globale Änderung von Selbstbewertung und Motivation, die auf den Einfluß der Noten (und der Prüfung) zurückgeht. Sie kann in ihren theoretischen Teilen auf mehrere unterschiedliche Ansätze zurückgreifen, die den Zusammenhang von Fremdbewertungen und den abhängigen Variablen dem Grunde nach wie auch im Hinblick auf bestimmte Bedingungen dieses Zusammenhangs analysieren, die auf personaler oder situativer Ebene bestehen können. Auch empirische Untersuchungen, die dem gewählten Ansatz und den konkreten Gegenstandsbedingungen so weit gleichen, daß sie von daher übertragbar sind, geben Aufschluß über mögliche Wirkungen. In dem ergänzenden empirischen Teil wird über die Ergebnisse der eigenen Untersuchung berichtet, die in verschiedenartigen Auswertungsansätzen gewonnen werden.

Eine Fortführung der Wirkungsanalyse bildet die Stufenanalyse (III.III.) mit dem theoretischen und empirischen Versuch, den globalen Prozeß der Wirkung der Noten wie der Prüfung in vermittelnden oder zwischengeschalteten Schritten zu erfassen. Die Differenzierung dient sowohl dem Aufzeigen möglicher Quellen der Verfälschung der globalen Wirkung als auch der Bestimmung von vermittelnden Prozessen, die den abhängigen Variablen vor- und den unabhängigen Variablen nachgeschaltet gedacht werden. Als solche vermittelnden Schritte werden bearbeitet: das Anspruchsniveau der Prüflinge; die erlebte Rückmeldung durch die Prüfung; die Struktur der subjektiven Notenskalen; die erlebte Selbstverantwortlichkeit für das Ergebnis der Prüfung.

Die theoretische und empirische Analyse vorbereitend beschäftigt sich der erste der folgenden beiden Abschnitte mit dem Verhältnis von Theorie und Empirie und daran anschließend mit dem Problem der Über-

tragung von Theorien, Begriffen und Ergebnissen von empirischen Untersuchungen auf das anstehende Problem. Beides ist für die richtige Einordnung der "theoretischen" Teile der jeweiligen Kapitel grundlegend. Das methodische Vorgehen der eigenen empirischen Untersuchung wird, sofern es für das Verständnis der empirischen Auswertungen und der Folgerungen daraus notwendig ist, in einem zweiten Abschnitt geschildert.

Bedeutung von Theorie und Empirie

Im strengen Sinn einer empirisch kontrollierten Erkenntnisgewinnung ist die Theorie in bezug auf die empirische Tätigkeit erkenntnisleitend. Sie determiniert diese Tätigkeit zwar nicht vollständig, stellt aber deren Grundlage dar, indem sie über die Ableitung von Hypothesen, die in empirischer Tätigkeit überprüft und als geprüfte an die Theorie rückvermittelt werden, einen Rahmen setzt. Wenn in der vorliegenden Untersuchung eine Theorie als solche markiert werden soll, dann eignet sich dafür die - nicht allein psychologische - Funktionentheorie des Kapitels II. Ihr jedoch fehlen konkrete Hypothesen.

Fällt die allgemeine Praxis der psychologischen Grundlagenforschung schon weit hinter dieses normative Modell der Erkenntnisgewinnung zurück, so verliert es für die vorliegende Fragestellung, die sich als angewandte versteht, auch seine normative Verbindlichkeit. Im analytischen Teil der Arbeit hat die psychologische Theorie den Stellenwert eines Mittels der Analyse und ist deshalb der Empirie eher gleichgestellt. In dieser Funktion umfaßt das, was im theoretischen Teil aufgearbeitet wird, neben eigentlichen Theorien, die in der gegenwärtigen akademischen Psychologie ohnehin ein Kümmerdasein führen, auch isolierte Begriffstraditionen und nur locker von Begriffen eingebettete empirische Untersuchungen. All das dient zur Gewinnung von

Erkenntnis, die auf das anstehende Problem übertragen wird. (Die pädagogische Psychologie insgesamt lebt in weiten Bereichen von diesen Übertragungen.) Eine Erkenntnisgewinnung durch Übertragung ist aus zwei Gründen für angewandte Fragestellungen notwendig und sinnvoll.

1. Für die Gegenstände einer solchen angewandten Fragestellung fehlt eine eigenständige Theorie, was in deren Komplexität und Wandelbarkeit seinen nicht zufälligen Grund hat.

2. Auch die Möglichkeit eigenständiger empirischer Erkenntnisgewinnung ist meist beschränkt und deshalb durch Übertragung von Erkenntnissen zu ergänzen.

Die Übertragung von Theorien und Begriffen läßt konkurrierende Ansätze der Psychologie nicht nur zu, sondern sogar wünschbar erscheinen, weil dadurch die gegenstandsangemessene Komplexität eher gesteigert wird. Dabei muß allerdings das verständliche Bedürfnis nach Klarheit (in der Entscheidung für eine Theorie) oder nach Integration unbefriedigt bleiben, weil das in diesem Zusammenhang kaum zu leisten und im Hinblick auf die Anwendung weder fruchtbar noch (in jedem Fall) sinnvoll ist. Das einheitsstiftende Element ist nicht die herausgefilterte theoretische und begriffliche Struktur, sondern der konkrete Gegenstand der praktischen Fragestellung.

Es ist nun kurz zu begründen, wodurch eine Übertragung psychologischer Theorien und Begriffe ermöglicht wird. Ganz allgemein läßt sich die Anwendung kennzeichnen als die Interpretation einer Theorie (oder auch nur eines einzelnen Konstruktes) auf den vorliegenden Gegenstand. Die Interpretation ist jedoch nur in einem bestimmten Rahmen sinnvoll, der durch eine genauere Klärung des Vorgangs der Interpretation gewonnen werden kann. Explizite Regeln zur Grenzbestimmung für solche Übertragungen liegen nicht vor; ihr Fehlen wird gerade auch in der pädagogischen Psychologie deut-

lich. Die folgenden Vorüberlegungen sind nur ein erster Schritt.

Eine wesentliche Form der Übertragung besteht darin, daß eine Theorie als Ganzes oder bestimmte ihrer Begriffe auf einen konkreten Gegenstand direkt angewendet werden. Die Gefahr eines solchen Vorgehens wird deutlich, wenn diese Form der Theorieanwendung mit einer spezifischeren kontrastiert wird, die die Anwendung von der Übereinstimmung der für erfahrungswissenschaftliche Theorien und Begriffe zentralen Form des Empiriebezuges, der Operationalisierungen, mit den Eigenschaften des Gegenstandes der praktischen Fragestellung abhängig macht. Ohne Rücksicht auf die Wirklichkeitsverankerung einer Theorie kann deren Anwendung am konkreten Gegenstand vorbeiführen. Das ist um so weniger der Fall, um so mehr eine Theorie, was zu fordern ist, vielfältig und umfassend, d.h. mit unterschiedlichen Operationalisierungen, wirklichkeitsverankert ist.

Die rein von der Begrifflichkeit ausgehende Interpretation einer Theorie auf einen Gegenstand ist dort unproblematisch, wo die Theorie selbst kaum streng empirisch verankert ist. Dann allerdings ist der Erkenntnisgewinn durch Anwendung der Theorie gegenüber unmittelbar gegenstandsbezogener begrifflicher Tätigkeit nicht immer ausreichend zur Rechtfertigung der Anwendung.

Auf der anderen Seite ist auch eine Übertragung allein von zu einer Theorie gehörigen empirischen Untersuchungen möglich, wobei die zugeordneten Begriffe und Theorien außer acht gelassen werden. Viele Untersuchungen eklektischer Art legen ein solches Vorgehen nahe, weil die in ihnen verwendeten Konstrukte kaum über eine ad hoc erstellte Beschreibung der Operationalisierung hinausgehen. Für die Übertragung der Ergebnisse von isolierten einzelnen empirischen Untersuchungen gilt der Hinweis, der so einfach zu formulieren wie

schwer zu beachten ist, daß die konkreten Umstände, unter denen die Untersuchung stattfindet, der Beachtung bedürfen. Das ist dort besonders zentral, wo komplexe, institutionengebundene Prozesse Gegenstand sind.

Die Verwendung der Ergebnisse einzelner empirischer Untersuchungen stellt die Nahtstelle her zur eigenen empirischen Untersuchung, wobei deren eher gleichgeordnete Stellung aus diesem Übergang noch einmal deutlich wird. Der eigene Zugriff zur Empirie ist im Verhältnis zur Theorie nicht als Hypothesenprüfung zu begreifen. Es ist ein solches Verhältnis im einzelnen Fall nicht auszuschließen, aber dem Grunde nach stellen Theorie und Empirie unterschiedliche Zugriffsweisen auf denselben Gegenstand dar. Wenn dennoch regelmäßig der "theoretische" Teil dem empirischen vorangestellt ist, wird deutlich, daß die Gleichordnung nicht vollkommen ist, sondern zumindest die Festlegung des Begriffsgebrauchs, die Begründung der Operationalisierungen im empirischen Teil und auch ein allgemeines Vorverständnis vorgängig "theoretisch" zu leisten sind.

Methodisches Vorgehen der empirischen Untersuchung

Zum Verständnis der empirischen Untersuchung, die in der folgenden Darstellung eine zentrale Erkenntnisquelle ist, sind einige Hinweise auf das methodische Vorgehen notwendig, die eine Zusammenfassung entsprechender Stellen des Exkurses I darstellen. Der gesamte Exkurs kann allerdings zu einer an dieser Stelle angeordneten chronologischen Lektüre nicht empfohlen werden, weil er einige Ausführungen enthält, die erst verständlich sind, wenn das Kapitel III.I. zur Kenntnis gelangt ist.

Die empirische Untersuchung versteht sich als Er-

kundungsuntersuchung, in der eine Fülle von Aspekten erforscht werden soll, aus denen die Arbeit nur einen Ausschnitt herausgreift. Die Untersuchung bedient sich der Methode der schriftlichen Befragung, weil die als der Situation angemessen gesehen wird. Es wird vermutet, daß bei der Alternative der mündlichen Befragung Hindernisse zur Beteiligung vieler Prüflinge bestehen. Für die Beantwortermotivation ist die Situation der schriftlichen Befragung ausreichend, weil die Prüflinge "relevante" Fragen vorfinden.

Die Untersuchung ist als prae-post-design angelegt, um sicherzustellen, daß nach der Prüfung beobachtete Zusammenhänge auf den Einfluß der Prüfung (und der Noten) zurückgeführt werden können. Das zwischenzeitliche Geschehen zwischen den beiden Befragungen jedoch läßt, weil unkontrolliert, auch andere Deutungen zu.

In den Fragebögen werden geschlossene Fragen verwendet. Ein großer Teil der Fragen, so auch die thematischen Fragen nach der Motivation, sind mit vierstufigen Antwortskalen versehen, die polar oder vollständig verbal verankert sind. Die Vorgabe weniger Antwortstufen soll den Befragten das Urteil erleichtern; das Fehlen einer mittleren Stufe soll bei der geringen Stufenzahl eine Massierung der Antworten auf die mittlere Stufe verhindern. Die Fragen sind in dem Sinne als direkt zu bezeichnen, als Konstrukte des Erlebens und Verhaltens unmittelbar angesprochen werden. Sie weisen deshalb keinen den Befragten verborgenen Inhalt auf. Die Gefahr solcher Fragen besteht darin, daß bestimmte Stereotypen in Form von Einstellungen und gezielte Verfälschungen der Antworten möglich sind. Ihr Vorteil liegt u.a. in der für die Betroffenen besseren "Einsehbarkeit" der mit ihnen gewonnenen Ergebnisse.

Die Fragebogenuntersuchung wurde mit zwei Vorformen vorbereitet, die im Fach Psychologie an der Univer-

sität Münster bei Prüflingen des Vordiploms und des Hauptdiploms erprobt wurden. Die erste Vorform beinhaltete im wesentlichen noch offene Fragen. Nur ein Teil der thematischen Fragen war weitgehend der Endform des Fragebogens ähnlich. Die zweite Vorform diente allein der Kontrolle der Vollständigkeit der Antworten und stimmte weitgehend mit der Endform überein.

Die Endform, deren in dieser Untersuchung ausgewertete Fragen im Anhang I aufgeführt sind, besteht aus einem Prüfungsvorbereitungsfragebogen (im folgenden: PVF) und einem Prüfungsfolgenfragebogen (im folgenden: PFF). Der PVF wurde zwei Wochen vor den mündlichen Prüfungen verteilt, der PFF sofort nach den mündlichen Prüfungen.

Die in dieser Untersuchung ausgewertete Gruppe von Befragten besteht ausschließlich aus Prüflingen des Vordiploms und des Hauptdiploms im Fach Psychologie am Psychologischen Institut der Universität Münster. Im Vordiplom (im folgenden: VD) wurden über drei Prüfungstermine (1970/1971) verteilt 72 Prüflinge befragt; 59,7 % der Fragebögen sind auswertbar. Im Hauptdiplom (im folgenden: HD) wurden in denselben Prüfungsterminen 50 Prüflinge befragt; 58 % der Fragebögen sind auswertbar. In einigen Auswertungen werden zusätzlich, weil die Fragenformulierung fast identisch ist, auch die Ergebnisse der Voruntersuchungen mit einbezogen; das wird jeweils deutlich mit dem Hinweis "Weite Stichprobe" herausgestellt.

Bei bestimmten Fragen sind Auslassungen oder unklare Antworten durch Einsetzungen korrigierbar. Die Noten werden direkt aus den Prüfungsakten übernommen. Die Verarbeitung der Daten erfolgt mit Hilfe verschiedener Programmsysteme und eigener Programme am Rechenzentrum der Universität Münster. Es werden bevorzugt parametrische Verfahren angewendet, teil-

weise werden auch verschiedene Verfahren parallel benutzt. Für die Verwendung parametrischer Verfahren sprechen die weitgehende Normalverteilung der Antworten und die Unersetzbarkeit der Verfahren.

Eine Kontrolle der Auswirkungen der Selbstselektion ergibt im VD und HD keinen Unterschied von Beantwortern und Nichtbeantwortern nach Geschlecht, Alter und gewählter Prüfungsordnung. Im VD unterscheiden sich die beiden Gruppen auch nicht gravierend in den Noten, die sie in den Prüfungen erhalten. Im HD dagegen sind die Beantworter eindeutig in den Noten der Einzelfächer und den Gesamtnoten den Nichtbeantwortern überlegen. Das schränkt die Verallgemeinerbarkeit der Ergebnisse für das HD schon dann ein, wenn sie auf die Gesamtgruppe der Prüflinge hin erfolgen soll.

Korrelationen der thematischen Fragen zwischen PVF und PFF ergeben, daß die Antworten als recht zuverlässig betrachtet werden können. Das bestätigt auch die Korrelation zwischen errechneten und beantworteten Gesamtwerten der Selbstbewertung. Die Mittelwerte aller einzelfachbezogenen Antworten bei vierstufigen Antwortskalen zeigen, daß die Beantwortung gegenstandsspezifisch erfolgt und nicht wesentlich von der Stellung der Frage im Rahmen des Fragebogens bestimmt zu sein scheint. Gleichheiten der Mittelwerte bei VD und HD aber schließen eine Mitwirkung personenspezifischen Gebrauchs der Antwortskalen nicht aus.

III.I.: Voraussetzungsanalyse

Übersicht: Das Kapitel gliedert sich in 5 Teile. Die ersten beiden Teile gelten den unabhängigen Variablen Prüfung und Noten; die übrigen drei den abhängigen Variablen Selbstbewertung und Motivation.

Der Teil A beschäftigt sich mit der Prüfung. Im Hinblick auf die konkret untersuchten Prüfungen am Psychologischen Institut der Universität Münster schil-

dert er die Umstände der Prüfung, die Grundlage möglicher Übertragung der Ergebnisse sind. Zugleich ist aus den Umständen heraus, die die Komplexität der erhobenen Daten verursachen, zu bestimmen, auf welchen Ebenen die Auswertung der Daten erfolgen soll. Ferner ist das bisher unerledigte Problem aufzunehmen, ob im zwischenzeitlichen Geschehen zwischen PVF und PFF nur die Prüfung als Erklärung von Veränderungen herangezogen werden kann.

Der Teil B gilt den Noten, die als institutionell vorgesehenes Mittel der Rückmeldung die primäre unabhängige Variable darstellen. Eine Auswertung der in Fülle vorliegenden Literatur zu Noten wird mit Ergebnissen kombiniert, die aus den im Rahmen dieser Untersuchung erteilten Noten gewonnen werden.

Der Teil C wendet sich der zentralen abhängigen Variablen, der Selbstbewertung zu. Da in der deutschsprachigen Literatur es an zusammenfassenden Darstellungen fehlt, soll der Erkenntnisstand in systematischer Weise berichtet werden. Die Begründung von Inhalt und Form der gestellten Fragen erfolgt auf dem Hintergrund der begrifflichen Klärungsversuche und von Ergebnissen empirischer Untersuchungen.

Der Teil D zur Motivation ist wesentlich kürzer gehalten, weil eine ausführliche theoretische Auseinandersetzung angesichts der Geschlossenheit vorhandener Theorien kaum zu leisten ist und eine eigenständige Tradition empirischer Untersuchungen mit ähnlichen Fragen nicht besteht. Auch in bezug auf die Selbstbewertung wird der motivationale Zustand der Prüflinge vor der Prüfung erörtert, um die Verwendung der Daten des PVF als Referenzdaten abzusichern.

Die nur kurze Vorklärung des Zueinanders von Selbstbewertung und Motivation in den Teilen C und D wird im abschließenden Teil E durch die empirische Auswertung der Daten ergänzt. Der Teil beinhaltet

darüber hinaus eine Kennzeichnung der Eigentümlichkeiten des Antwortverhaltens, die zusammen mit der Analyse der Beziehungen zwischen den Antworten eine Abschätzung der Validität der angesprochenen Prozesse ermöglichen soll.

III.I.A.: Prüfungen als Hintergrund und Ursache

Prüfungsbedingungen am Psychologischen Institut der Universität Münster

Die genaue Kenntnis der Prüfungsbedingungen ist notwendig, um die Ergebnisse der empirischen Untersuchung auf andere Fächer und andere Institute zu verallgemeinern (s. Exkurs I).

Die Rahmenbedingungen, die die Prüfungen am Psychologischen Institut der Universität Münster zur Zeit der empirischen Untersuchung kennzeichnen, sind auf zwei Ebenen determiniert. Auf der formellen Ebene werden sie gebildet aus den Regelungen der Prüfungsordnung; auf der informellen werden sie gebildet aus den empirischen Verwirklichungen der Prüfungsordnung, die von der Auslegung durch die einzelnen Prüfer bis zur Tradition des Prüfungswesens dieses Instituts reichen. Nur die allgemeine Tradition der Auslegung wird in diesem Teil als Rahmenbedingung erfaßt, die Auslegung durch die einzelnen Prüfer wird, wenn überhaupt, erst aus den Ergebnissen deutlich. Zur Analyse der Auslegung wären u.a. empirische Untersuchungen notwendig. Die hier mitgeteilten Beobachtungen sind mehr oder weniger subjektive Einschätzungen.

Im Zeitraum der Untersuchung vollzieht sich die Umstellung von einer alten Prüfungsordnung (vom 1.5.1956; im folgenden: APO) zu einer neuen Prüfungsordnung (vom 27.10.1970; im folgenden: NPO). Diese Umstellung war vor der Planung der Untersuchung nicht absehbar. Die faktischen Änderungen sind jedoch, weni-

ger als die Änderungen der Bestimmungen der Prüfungsordnungen andeuten, nicht so gravierend, daß nicht eine zusammenfassende Auswertung der Ergebnisse gerechtfertigt erscheint. Dennoch soll an einigen Stellen der Einfluß der unterschiedlichen Prüfungsordnungen, die zur Zeit der Untersuchung alternativ wählbar sind, kontrolliert werden.

Die Prüfungen setzen sich nach beiden Prüfungsordnungen aus einer Vorprüfung und einer Hauptprüfung zusammen: dem Vordiplom und dem Hauptdiplom.

Im Vordiplom sieht die APO die Ablegung der Prüfung frühestens nach dem 5. Semester vor. Zur Zulassung ist das Vorlegen einer schriftlichen Hausarbeit Bedingung. Die Prüfung besteht dann aus mehreren mündlichen Prüfungen von jeweils dreißigminütiger Dauer, die von einem Prüfer, den ein Protokollant begleitet, abgenommen werden. Die mündlichen Prüfungen werden in den folgenden Einzelfächern verlangt: Ausdruckskunde; Allgemeine Psychologie; Entwicklungspsychologie; Charakterkunde; Physiologie; Philosophie; Biologie oder Humangenetik (Humangenetik hat sich fast ohne Ausnahme durchgesetzt).

Die Skala zur Bewertung der Leistungen in der Hausarbeit und den mündlichen Prüfungen umfaßt die Ziffernnoten von 1 - 5. Aus den Einzelnoten wird durch Gewichtung und Mittelung die Gesamtnote derart gebildet, daß die Hausarbeit doppelt gewichtet wird, die Nebenfächer Philosophie und Humangenetik nur halb gewichtet werden. Es gilt folgende Zuordnung der Gesamtnoten zu den arithmetischen Mittelwerten:

Gesamtnote	Mittelwert
1	1 - 1,5
2	1,6 - 2,5
3	2,6 - 3,5
4	3,6 - 4,3

Demgegenüber ist in der NPO eine Zulassung bereits nach 4 Semestern vorgesehen. Die Hausarbeit ist gestrichen; die mündlichen Prüfungen bleiben erhalten. Das Fach Ausdruckskunde wird durch das Fach Methodenlehre ersetzt; das Fach Charakterkunde wird in Persönlichkeitspsychologie umbenannt (im folgenden wird deshalb nur noch zur Bezeichnung "Persönlichkeitspsychologie" verwendet). Die Nebenfächer werden weitgehend frei wählbar, bleiben im Zeitraum der Untersuchung aber fast ausschließlich die aus der APO bekannten Fächer.

Die Skala zur Bewertung bleibt dieselbe. Die Bildung der Gesamtnote erfolgt durch doppelte Gewichtung der psychologischen Fächer und Physiologie und einfache Gewichtung der beiden Nebenfächer. Die Zuordnung der Gesamtnote zu den Mittelwerten entspricht der Regelung in der APO.

Das Hauptdiplom ist nach der APO frühestens 3 Semester nach dem VD abzulegen. Es sieht eine vierstündige Klausur in einem theoretischen Bereich und eine Fallklausur von unterschiedlicher Dauer im klinisch-diagnostischen Bereich vor.

Die mündlichen Prüfungen gleichen in der Form denen des VD und erstrecken sich auf die folgenden Fächer: Psychologische Diagnostik; Tiefenpsychologie; Pädagogische Psychologie; Sozialpsychologie; Angewandte Psychologie und Psychopathologie.

Für die Gesamtnote werden die Klausuren doppelt gewichtet. Im übrigen sind Notenskala und Zuordnung der Gesamtnoten zu den Mittelwerten wie im VD geregelt.

Demgegenüber schreibt die NPO eine Dauer von 4 Semestern Studium nach dem VD vor. Eine schriftliche Arbeit muß zusätzlich als Zulassungsarbeit vorgelegt werden. Die theoretische Klausur fällt fort, eine vier- bis achtstündige praktische Klausur kann aus allen Gebieten der Fächer der mündlichen Prüfungen ge-

wählt werden. Das Fach Psychopathologie fällt in der mündlichen Prüfung weg (es wird deshalb in der Auswertung insgesamt nicht berücksichtigt); an die Stelle des Faches Tiefenpsychologie tritt das Fach Klinische Psychologie (der Inhalt aber bedarf keiner Änderung; im folgenden wird deshalb nur noch die Bezeichnung "Klinische Psychologie" verwendet).

Bei der Gesamtnotenbildung wird das Ergebnis der Arbeit doppelt gewichtet. Notenskala und Zuordnung der Gesamtnoten ändern sich nicht.

Im Zentrum aller Prüfungen stehen also die mündlichen Prüfungen in den einzelnen Fächern. Dementsprechend konzentriert sich die empirische Untersuchung neben den Gesamtnoten auf diese mündlichen Prüfungen, wie schon aus dem Ablauf der Befragung und der Gestaltung des Fragebogens deutlich wird. Die Hausarbeit und die Klausuren finden nur mittelbar durch ihre Verrechnung in der Gesamtnote Berücksichtigung. Die große Bedeutung der mündlichen Prüfungen eignet trotz aller sonstigen Unterschiede allen damaligen Prüfungsordnungen im Fach Psychologie in der BRD (GASCH 1971; HASEMANN 1970a).

Auf der informellen Ebene sind die Prüfungen am Psychologischen Institut der Universität Münster im Untersuchungszeitraum durch die Einflüsse reformerischer Rationalisierung und Liberalisierung gekennzeichnet. Es ist weitgehend üblich, den Studenten explizite Information über die Anforderungen zur Verfügung zu stellen. Mit Zustimmung des Prüflings kann eine begrenzte Zahl von Zuhörern an den mündlichen Prüfungen teilnehmen (Öffentlichkeit). Viele Prüfer legen den Ablauf der mündlichen Prüfung über die Bestimmungen der Prüfungsordnung hinausgehend fest. Doch zeigen die ausgewerteten Tonbandprotokolle mündlicher Prüfungen (SCHLIEHE & SEEGER 1974), daß

die im wesentlichen syntaktisch erfaßten verbalen Interaktionen von Prüfer zu Prüfer weitgehend unterschiedlich sind.

Im VD wird ein wesentlicher Anteil von mündlichen Prüfungen durch Assistenten durchgeführt, im HD sind an den mündlichen Prüfungen mehr die Professoren beteiligt. In vielen Fächern besteht die Möglichkeit, mindestens zwischen zwei Prüfern zu wählen. Darin unterscheiden sich allerdings nicht nur die einzelnen Prüfungsfächer, sondern auch die einzelnen Prüfungstermine.

An dieser Stelle sollen kurz zusätzlich einige Merkmale der Studenten und des Studiums berichtet werden, die gleichfalls bei der Übertragung der Ergebnisse von Bedeutung sein können.

Das Studium zum VD weist einen geringen Anteil von Vorlesungen und einen großen Anteil von Kursen und Übungen auf. In den naturwissenschaftlichen Nebenfächern allerdings dominieren die Vorlesungen. Es ist eine Reihe von Veranstaltungen zu absolvieren, die Voraussetzungen zur Prüfungszulassung darstellen (Experimentelle Übungen, Statistikkurse). Aber auch darüber hinaus ist die Wahlfreiheit von Veranstaltungen weitgehend eingeschränkt.

Das Studium zum HD weist kaum noch Vorlesungen auf; es dominiert der klinisch-diagnostische Bereich mit intensiver Ausbildung. Das Veranstaltungsangebot in den Fächern Sozialpsychologie und Angewandte Psychologie ist weniger umfangreich, das in Pädagogischer Psychologie gering.

Die Prüflinge, die einen ausgefüllten Prüfungsvorbereitungsfragebogen zurückgeschickt haben, sind durch die folgenden Daten zu kennzeichnen: Die Prüflinge des VD absolvieren die Prüfung im Durchschnitt nach 6,9 Semestern (wie die folgenden Angaben: arith-

metisches Mittel). Für das Studium nach der APO wird annähernd 1 Semester durchschnittlich mehr gebraucht als für das Studium nach der NPO (7,3 gegenüber 6,4 Semestern). Das HD wird nach 10,2 Semestern absolviert; davon sind 3,8 Semester seit dem VD studiert worden.

Die Mehrzahl der Prüflinge bereitet sich wenigstens in einem Fach in Vorbereitungsgruppen auf die Prüfung vor. Mindestens die Hälfte der Prüflinge macht Gebrauch von der Möglichkeit, sich in öffentlichen mündlichen Prüfungen über die Anforderungen der Prüfung und deren Durchführung zu informieren. Von wenigen Ausnahmen abgesehen haben die Prüflinge des HD ihr gesamtes Studium in Münster verbracht.

Ebenen der Auswertung für die Fächer der mündlichen Prüfungen

Die Auswertung für die Einzelfächer der mündlichen Prüfungen ist durch die Komplexität der Daten und der Bedingungen erschwert. Da die Einzelfächer aber, wie angemerkt, im Zentrum der empirischen Untersuchung (und der Prüfung insgesamt) stehen, soll die Auswertung detailliert begründet werden. Das geschieht im Exkurs II, dessen Ergebnisse die folgende Darstellung zusammenfaßt.

Für die Antworten zu den auf einzelne mündliche Prüfungen gerichteten Fragen (Fächerfragen) sind verschiedenartige Auswertungen möglich. Diese Auswertungen erfassen unterschiedliche Einflußquellen. Auf Grund einer Analyse möglicher Einflußquellen und einer Gliederung von Auswertungsansätzen werden drei Ebenen der Auswertung ausgewählt.

Die Auswahl geht davon aus, daß die Trennung der Einflüsse der einzelnen Prüfungssituation und der Rahmenbedingungen (Prüfungsfach, Prüfer und Prüfungsart) am besten durch eine Auswertung auf der Ebene der einzelnen "Fach-Prüfer-Kombination" gewährleistet werden kann.

Als "Fach-Prüfer-Kombination" werden die Prüfungen bezeichnet, die ein Prüfer in einem Prüfungsfach abnimmt.

Eine Zusammenfassung aller Einflußquellen ist bei der Auswertung auf der Ebene "Über-alle-Fächer" möglich, die die einzelnen Prüfungen als Elemente der Erhebung verwendet und die Daten über alle einzelnen Prüfungen hinweg auswertet.

Nur der Einfluß der Rahmenbedingungen wird auf der Ebene "Mittelwerte der Fach-Prüfer-Kombinationen" erfaßt, auf der die Mittelwerte, die über die Prüfungen einer "Fach-Prüfer-Kombination" hinweg berechnet werden, als neue Ausgangsdaten dienen.

Die Interpretation auf allen drei Ebenen wird dadurch erschwert, daß die Prüflinge den einzelnen "Fach-Prüfer-Kombinationen" nicht nach Zufall zugeordnet sind, sondern sich zumindest teilweise die Prüfer selbst wählen können. Auch Prüfer und Prüfungsfach sind miteinander konfundiert, weil bestimmte Prüfer nur in bestimmten Prüfungsfächern prüfen. Gleichzeitig sind Prüfungsart (VD und HD) und Prüfer miteinander konfundiert, indem in den beiden Prüfungsarten nicht dieselben Prüfer prüfen.

Ein weiteres Problem der Auswertung auf den beiden Ebenen "Über-alle-Fächer" und "Mittelwerte der Fach-Prüfer-Kombinationen" besteht darin, daß die einzelnen Werte teilweise voneinander abhängig sind und deshalb die Voraussetzungen von inferenzstatistischen Verfahren nicht voll erfüllt sind. Die Signifikanzangaben für diese Auswertungen werden deshalb durch ein (?) gekennzeichnet.

Eingrenzung der Wirkung der Prüfungen

Die empirische Erhebung versucht als prae-post-design, die Wirkung auf die Prüfung und die Noten zu lokalisieren. Die Prüfung in diesem Sinne umfaßt die über mehrere Wochen verteilte Durchführung der mündli-

chen Einzelfachprüfungen und Klausuren (nur HD), während die Zulassungsarbeit in den meisten Fällen vor dem PVF fertiggestellt ist und deren Note bekannt sein kann. Bei Änderungen zwischen PVF und PFF sind auch dann, wenn ihr Zusammenhang mit den Noten nachweisbar ist, zwei Fragen zu stellen:

1. Haben nicht die wesentlichen Einflüsse der Prüfung schon vor dem PVF stattgefunden?

2. Gehen die Einflüsse wirklich auf die Noten (oder die Prüfungen) zurück?

Das durch die erste Frage aufgeworfene Problem entsteht daraus, daß sich der Einfluß der Prüfung nicht auf die Zeit zwischen den beiden Fragebögen beschränkt. Schon die Zeit vor der ersten Befragung ist als Prüfungsvorbereitung (s. Kapitel II) durch die Prüfung beeinflußt. Auch in dieser Zeit erfolgen Rückmeldungen und sind Einflüsse auf Motivation und Selbstbewertung möglich. So zeigt die Befragung von ALVES & LODDERS (1974), daß im VD ein wesentlicher Anteil aller Strukturbildungen des Studiums in der Prüfungsvorbereitung erfolgt. Dementsprechend wäre mit der Absicht, den Einfluß der Prüfung zu untersuchen, dieser gesamte Zeitraum samt Vorbereitung in die Untersuchung einzubeziehen. Die zitierte Befragung weist jedoch auch nach, daß der Anteil des in der Prüfungsvorbereitung Erarbeiteten in der Vorbereitung auf das HD sehr viel geringer wird, weil sich die Erarbeitung ins Studium verlagert. Damit scheint erwiesen, daß studienorganisatorische Bedingungen unabhängig von der Prüfung eine Erarbeitung des Stoffes ermöglichen können und diese nicht an die Prüfungsvorbereitung gebunden ist. Die Rückmeldung, die im Rahmen der Prüfungsvorbereitung erfolgt und sich auf Motivation und Selbstkonzept auswirken kann, ist prinzipiell auch (und besser ?) im Studium möglich. Der PVF kann deshalb als Ausgangsmessung verwendet werden. (Diese Argumen-

tation betrifft noch nicht mögliche Verzerrungen der Antworten im PVF im Hinblick auf die kommenden Prüfungen. Dazu nimmt Teil D u.a. Stellung.)

Die zweite Frage äußert Zweifel an der Kausalität beobachteter Wirkungen. Vor jeder einzelnen mündlichen Prüfung findet noch eine intensive Auseinandersetzung mit den Inhalten statt, die Prüfungen selbst enthalten teilweise viele informelle Rückmeldungen verbaler und nonverbaler Art. Wenn z.B. vermutet wird, daß Wirkungen auf die Noten zurückzuführen sind, und dies mit entsprechenden empirischen Kontingenzen belegt wird, dann ist durchaus möglich, daß die eigentlichen Bewirkungen von solchen Prozessen ausgehen, die mit den Noten konfundiert sind. Dieses Problem ist in der vorliegenden Untersuchung nur unter Hinzunahme von theoretischen Überlegungen und Ergebnissen anderer empirischer Untersuchungen angehbar. In dieser Hinsicht ist die Untersuchung mit querschnitthaften Studienerfolgsuntersuchungen vergleichbar; nur komplex angelegte Verlaufsuntersuchungen, die aber den Nachteil mit sich brächten, den wirklichen Prozeß durch vielfache Eingriffe der unerforschten Wirklichkeit immer mehr zu entfremden, oder Experimente könnten weitergehende empirische Belege beibringen. (Gegen die Wirkung informeller Rückmeldung in Prüfungen allerdings spricht, daß die Antworten im PFF keine Zusammenhänge mit den Daten der Untersuchung von SCHLIEHE & SEEGER (1974) zeigen, die in einem Teil der mündlichen Prüfungen im Untersuchungszeitraum die Interaktionen auch im Hinblick auf - allerdings formal verstandene - Rückmeldungen auswertet.)

Zusammenfassung (III.I.A.)

Die Prüfungen am Psychologischen Institut der Universität Münster bestehen aus einer Vordiplomprüfung

(im folgenden: VD) und einer Hauptdiplomprüfung (im folgenden: HD). Im Zeitraum der Untersuchung findet eine Umstellung von einer alten zu einer neuen Prüfungsordnung (im folgenden: APO und NPO) statt. Im Zentrum der Prüfungen stehen bei beiden Prüfungsordnungen mündliche Prüfungen in verschiedenen Fächern. Die Unterschiede zwischen den beiden Prüfungsordnungen sind insgesamt gering, so daß die Auswertung zusammengefaßt erfolgt. Die Durchführung der mündlichen Prüfungen ist durch reformerische Rationalisierung gekennzeichnet. Das Studium selbst ist weitgehend in Übungen und Seminaren organisiert.

Die Antworten zu den Einzelfächern der mündlichen Prüfungen werden auf den folgenden Ebenen ausgewertet: "Fach-Prüfer-Kombination" (Prüfungen, die ein Prüfer in einem Prüfungsfach abnimmt); "Über-alle-Fächer" (die einzelnen Prüfungen und nicht die Prüflinge werden als Elemente der Erhebung betrachtet); "Mittelwerte der Fach-Prüfer-Kombinationen" (aus den Daten der ersten Ebene wird ein Mittelwert der Antworten aller Prüflinge in der Kombination berechnet). Die Auswertungen auf den Ebenen "Über-alle-Fächer" und "Mittelwerte der Fach-Prüfer-Kombinationen" verwenden teilweise voneinander abhängige Werte. Die Angaben für statistische Inferenzen bei diesen Auswertungen werden deshalb mit einem (?) versehen.

Einflüsse der Prüfungsvorbereitung auf die thematischen Variablen werden nicht als Störbedingungen betrachtet, sondern als Ergebnis von Prozessen, die ein vernünftig strukturiertes Studium auch unabhängig von der Prüfung müßte bereitstellen können. Ein Problem der Untersuchung bleibt aber, daß die Kausalität der Prüfungen (und Noten) für bestimmte Ergebnisse nur in einer Verlaufsuntersuchung eindeutig verifiziert werden könnte.

III.I.B.: Eigenschaften der Noten

Die Noten sind das einzige institutionell vorgesehene Medium der Mitteilung des Prüfungserfolgs an die Prüflinge (Kapitel II.4.). Aus einer theoretischen und empirischen Analyse der "technischen" Eigenschaften der Noten ist Aufschluß darüber zu erlangen, wie weit die Noten sich als Mittel der Rückmeldung eignen. Die Analyse kann von einer Vielzahl von einzelnen empirisch orientierten Untersuchungen ausgehen; zusätzlich können Überprüfungen direkt an den Noten dieser Untersuchung vorgenommen werden. Dazu werden auch die Ergebnisse einer Untersuchung am vollständigen Notenmaterial der drei Prüfungstermine und an weiterem Material aus vorhergehenden Jahren herangezogen (WOTHKE 1974).

Die Darstellung soli Vorhersagen möglicher Wirkungen der Noten herausarbeiten oder zumindest Material für die nachträgliche Erklärung bestimmter Wirkungen bereitstellen. Die Eigenschaften der Noten sind mit dem Prozeß der Wirkung nicht allein funktional vermittelt, sondern es ist auch davon auszugehen, daß bestimmte Eigenschaften den Prüflingen nicht verborgen bleiben und so direkt in den subjektiven Prozeß der Verarbeitung der mit den Noten übermittelten Information eingreifen.

Übersicht: Grundlage der Aussagekraft von Noten sind deren skalometrische Eigenschaften und die Eigenarten der empirischen Notenverteilungen. Sie sind ausschlaggebend für den Informationsgehalt der Noten, die Möglichkeiten und Implikationen der Bildung von Gesamtnoten und die Transformation von Noten (Teil 1).

Der überwiegende Teil vorliegender empirischer Untersuchungen zu Noten gilt den aus der Testpsychologie stammenden Eigenschaften wie Zuverlässigkeit und Validität. Dabei sind nicht nur die Produkte, sondern

auch die Prozesse, auf die die Noten zurückverweisen, in die Betrachtung einzubeziehen (Teil 2).

Der Teil 3 versucht, die dabei gewonnenen Erkenntnisse teilweise an den Noten dieser Untersuchung zu überprüfen. Es ist zwar keine direkte Bestimmung der Zuverlässigkeit und Validität möglich; statt dessen wird die Wirkung einzelner Bedingungen wie der Prüfungsart, der Prüfungsumstände, der Prüfer und auch Eigenschaften der Prüflinge auf die Noten ermittelt.

Abschließend soll kurz (Teil 4) eine Integration über die Beantwortung der Frage erfolgen, ob die Noten zur Verwirklichung der Rückmeldefunktion geeignet erscheinen.

III.I.B.1.: Statistische und skalometrische Eigenschaften der Noten

Der informative Gehalt von Noten, der sowohl bewußte als auch funktionale Grundlage der Einwirkung auf die Selbstbewertung und die Motivation sein kann, ist über die Struktur der Notenskala und über die Verteilungseigenschaften der Noten zu bestimmen. Skalenstruktur und Verteilungseigenschaften haben Konsequenzen für die Bildung von Gesamtwerten und die statistische Umrechnung (Transformation) der Noten.

Struktur der Notenskala (Prüfer)

Eine Bestimmung der Struktur der Notenskala erfolgt zweckmäßigerweise durch Einordnung in die auf STEVENS zurückgehende Klassifizierung von Skalentypen (GUTJAHR 1974; PFANZAGL 1959): Nominalskala, Ordinalskala, Intervallskala und Rationalskala. In den Sozialwissenschaften ist die Zuordnung zu den axiomatisch klar geschiedenen Typen von Skalen jedoch dann ein Problem, wenn die Meßprozesse selbst nicht offen-

liegen, sondern wie bei den Noten nur als Ergebnis komplexer Urteilsprozesse gegeben sind, in denen die Urteiler das Instrument der Skalierung darstellen (ATHANASIADIS 1974). Die Zuordnung ist zwar insofern vereinfacht, als sie sich allein zwischen den beiden "mittleren" Typen der Ordinalskala und der Intervallskala zu entscheiden hat.

In der praktisch-pädagogischen Verwendung ist über die Zuordnung bereits entschieden: Dort werden Gesamtnoten durch arithmetische Mittelung berechnet, was die Voraussetzung hat, daß Noten auf einer Intervallskala gemessen werden. Eine arithmetische Bildung von Mittelwerten setzt in jedem Fall voraus, daß die Abstände zwischen den zu mittelnden Werten bekannt sind (auf einer Rationalskala); im Fall der Mittelung durch einfache arithmetische Mittelwertsbildung ist sogar vorausgesetzt, daß die Abstände zwischen den Notenstufen zahlenmäßig gleich sind. Eine gleichabständige Intervallskala als Grundlage der Noten wird auch angenommen, wenn lineare Transformationen Notenverteilungen besser vergleichbar machen sollen, wie dies ASCHERSLEBEN (1970) vorschlägt.

Vom skalometrischen Standpunkt aus muß die Zuordnung voreilig, weil unbegründet erscheinen. Doch leistet die Psychologische Methodenlehre wenig konstruktive Hilfe. Sie nimmt die Zuordnung gleichfalls eher "dogmatisch" vor, nur mit dem Unterschied, daß Noten als ordinalskaliert betrachtet werden. Eine solche Zuordnung findet sich häufig in statistischen Texten (z.B. CLAUSS & EBNER 1975), wo sie mit derselben Selbstverständlichkeit postuliert wird, wie sie in der Forschungspraxis übergangen wird, indem Methoden, die Intervallskalierung voraussetzen, angewendet werden.

Auch Arbeiten, die sich mit der praktischen Verwendung von Noten beschäftigen, billigen diesen nicht

mehr als ordinale Skalierung zu (GUTJAHR 1974; HASEMANN 1970b). Mit GUTJAHR (1974) und ATHANASIADIS (1974) kann die Bestimmung noch präzisiert werden, indem die Skala als homomorphe Ordinalskala gekennzeichnet wird, in der nicht die einzelnen Meßobjekte geordnet sind, sondern die Klassen, in die die Objekte zur Messung eingeordnet werden. (Bei GUTJAHR wird dies als Ordinalskala in Absetzung von der Rangskala bezeichnet.) Die Annahme, daß solche Messungen notwendigerweise nur die Voraussetzungen einer Ordinalskala erfüllen, wird durch SIXTL (1967) vom methodologischen Standpunkt psychologischer Skalierung aus nicht unbedingt bestätigt. Nach SIXTL sind Noten direkte Messungen, bei denen die Messung unmittelbar durch die Personen in der Art eines Ratings (Einschätzung) vorgenommen wird. Solche Messungen können durchaus intervallskaliert sein, wenn bei der Einschätzung z.B. die Anweisung erteilt wird, gleiche Abstände einzuhalten.

Antworten zu den damit aufgeworfenen Fragen sind nur zu erlangen, wenn die Zuordnung zu Skalentypen selbst Objekt empirischer Untersuchungen wird (KRISTOF 1969). Über eine globale Zuordnung hinausweisend wäre zu fragen, welche konkreten Auswirkungen die fehlerhafte Zuordnung hat: z.B. nach den Auswirkungen auf die Verwendung statistischer Methoden, die Intervallskalierung voraussetzen und auf Noten schon deshalb häufig angewendet werden, weil Verfahren für homomorphe Ordinalskalen fehlen (GUTJAHR 1974). Oder es wäre auch zu fragen nach der Verzerrung der übermittelten Information, wenn die Abnehmer der Information eine Intervallskalierung fälschlich annehmen.

Eine empirische Untersuchung der Struktur der Notenskala ist auf zwei grundsätzlichen Wegen üblich: auf einem direkten, der versucht, die Abstände zwischen den einzelnen Notenstufen zu ermitteln, und einem indirekten (problematischen) Vorgehen, das über

die Annahme der Normalverteilung arbeitet.

Direkte Versuche zur Bestimmung der Struktur von Notenskalen sind selten. Es sind zwei methodische Richtungen daran beteiligt: die Methode der SEU-Skalierung (subjective estimated utility) und der Paarvergleich nach THURSTONE. Beide Methoden wollen die subjektiven Abstände der einzelnen Notenstufen erfassen. Die Methode der Nutzenskalierung (SEU) stellt eine mehr funktionale Bestimmung der Abstände, die des Paarvergleichs eine mehr kognitiv-rationale dar. Die Nutzenskalierung ist bisher im Anschluß an BECKER & SIEGEL (1958; 1962) nur auf die Objekte von Noten (Studenten und Schüler) angewendet worden (s. Kapitel III.III). HERBST (1972) versucht als erster, den Ansatz auf Lehrer als Prüfer zu übertragen. Er findet bei den in Nützlichkeitsurteilen realisierten Skalen deutliche Abweichungen von der Gleichabständigkeit der Notenstufen. Die einzelnen Lehrer haben individuell sehr unterschiedliche Skalen; am häufigsten sind solche Skalen, die die größten Abstände zwischen den schlechten Noten aufweisen.

Das zweite Verfahren der Bestimmung von Abständen besteht darin, die einzelnen Abstände paarweise vergleichen zu lassen. Es scheint noch weniger als die SEU-Skalierung dem Prozeß der Notenvergabe funktional zu entsprechen. Bisher ist auch dafür nur eine Untersuchung bekannt, die Prüfer urteilen läßt (STONE & SINNETT 1968). Die in den USA durchgeführte Untersuchung unterteilt die Notenskala durch Hinzufügung von Plus- und Minusstufen in zwölf Intervalle für den Paarvergleich. Die ermittelten empirischen Notenskalen weisen eine logarithmische Beziehung zur äquidistanten Skala auf; die Abstände nehmen in Richtung der guten Noten zu; die Intervalle, die eine Notengrenze umfassen, sind phänomenal größer als die innerhalb einzelner Notenstufen; der Abstand von den Notenmitten zu

den Plusstufen ist größer als der zu den Minusstufen. Die Konsistenz der individuellen Urteile, auf denen diese Auswertungen aufbauen, ist gering, doch statistisch signifikant. Es werden methodenbedingt keine Versuche unternommen, individuelle Typen von Notenskalen zu unterscheiden.

Zunächst ist ein Befund deutlich herauszustellen: Die subjektiven Notenskalen von Prüfern, die täglich mit Noten umgehen, sind interindividuell sehr unterschiedlich; sie weichen deutlich von äquidistanten Skalen ab. Die Richtung der Abweichung ist allerdings je nach Methode verschieden; die SEU-Skalierung ergibt eher linksverschobene, der Paarvergleich eher rechtsverschobene Skalen. Es sind einander nicht ausschließende Erklärungen für die Unterschiede möglich. Der Unterschied könnte in den einbettenden Situationsfaktoren begründet sein, die Noten entweder mehr vom positiven Wert des Erfolgs her charakterisieren (USA) oder mehr den Mißerfolg (BRD) betonen. Die Ergebnisse von HERBST (1972) über Unterschiede zwischen den Lehrergruppen von zwei Schulen belegen die Bedeutung von Situationsfaktoren (Schulklima). Aber auch ein methodisches Artefakt ist dadurch nicht ausgeschlossen, daß STONE & SINNETT (1968) eine "künstliche" Aufteilung in Notenintervalle innerhalb von Notenstufen vornehmen. Die Vermutung bestätigt das andersartige Ergebnis der Anwendung des Paarvergleichs auf Studenten, die bei STONE (1969) mit den zwölf Intervallen wiederum größere Abstände zwischen den guten Noten realisieren, während CHANSKY (1969) für Paarvergleiche der ganzen Noten eindeutig zunehmend größere Abstände mit der Verschlechterung der Noten findet.

Mit aller Vorsicht wird man diesen Ergebnissen entnehmen, daß eine Gleichabständigkeit nicht vorausgesetzt werden kann. Die Auffassung der Notenskala als gleichabständige Intervallskala ist fragwürdig. Doch eine Antwort auf die zweite Frage nach den prak-

tischen Auswirkungen fälschlicher Verwendung als Intervallskala läßt der Erkenntnisstand nicht zu.

Die indirekte Analyse der Struktur der Notenskala nimmt den Weg über die Auswertung der Notenverteilungen. Dabei wird die folgende Argumentation verwendet: Die wirklichen Leistungen, die mit den Noten gemessen werden sollen, sind normalverteilt. Die Normalverteilung können Messungen nur dann wiedergeben, wenn die Messungen auf einer gleichabständigen Intervallskala erfolgen. Sind die Messungen empirisch normalverteilt und damit der wirklichen Verteilung der Leistungen isomorph, dann muß der Messung also eine solche Intervallskala zugrunde liegen. GUTJAHR (1974) weist mit Recht darauf hin, daß diese Argumentation problematisch ist. Jede Ermittlung von Verteilungen setzt schon die Annahme der Messung mindestens auf dem Niveau einer Intervallskala voraus. Auch die biologistische Grundannahme der Normalverteilung wirklicher Leistungen ist fragwürdig. Es ist deshalb sinnvoll, die Analyse der Notenverteilungen von der Ermittlung der Skalenstruktur der Notenskalen abzukoppeln.

Verteilungseigenschaften der Noten

Die Analyse der Verteilungseigenschaften von Noten stellt gewöhnlich eine Vermengung von präskriptivem und deskriptivem Vorgehen dar. Bestehende Verteilungen werden mit dem Modell der Normalverteilung verglichen. Das Modell der Normalverteilung wird dabei nicht nur als abstrakte Vergleichsgrundlage, sondern als praktisch zu verwirklichendes Ziel eingeführt. Als solches bedürfte es aber einer Rechtfertigung, die über eine implizite biologistische Grundannahme hinausgeht.

Normalverteilung als Vorschrift ist auch nicht

ohne immanente Probleme. Die Forderung nach Normalverteilung der Noten hat einen Höhepunkt in den 20er bis 40er Jahren zu verzeichnen (SMITH & DOBBIN 1960). Sie steht in enger Verbindung zum Aufschwung der Testverwendung und leitet von der Normalverteilung der Testergebnisse die der Noten ab (RINSLAND 1938). Im deutschsprachigen Bereich kommen solche Forderungen mit politisch bedingter Verspätung auf (LEHMANN 1955). Ausgehend von der festgelegten Stufenzahl der Noten wird verlangt, daß sich die empirischen Noten über diese Stufen normal verteilen. Die einzelnen Vorschriften kommen dabei zu deutlich unterschiedlichen Modellen, je nachdem, ob sie beispielsweise das Ziel haben, die mittlere Note in der Mitte einer Notenstufe anzusiedeln (EELLS 1930), die Varianz durch Gruppierung zu verringern (KELLEY 1950) oder aber eine 6-stufige Binomialverteilung zu erreichen (LEHMANN 1955). Man mußte auch erkennen, daß sich das Modell nur für große Gruppen eignet, während z.B. Schulklassen oder Prüflinge eines Prüfungstermins sogar Abweichungen davon zeigen müssen. Entsprechend wird es als praktische Anweisung nicht mehr empfohlen (STANLEY & HOPKINS 1972). Es soll deshalb ausschließlich in seiner deskriptiven Bedeutung verwendet werden. (Das präskriptive Normalverteilungsmodell gewinnt wieder an Bedeutung bei der nachträglichen Skalierung von Noten; s. den Abschnitt zur Transformation in diesem Teil.)

Empirische Notenverteilungen weisen meist deutliche Abweichungen von der Normalverteilung auf. Die Abweichungen betreffen Verschiebungen der Mittelwerte. Sie sind in Schulen eindeutig zu den besseren Noten hin verschoben (BURGER 1967; FALK 1962); dabei unterscheiden sich die einzelnen Schulfächer voneinander (KNOCHE 1969). Darüber hinaus finden sich Asymmetrien der Verteilungen (FLAMMER 1971). Die selbst schon in unterschiedlichen Schulsystemen gewonnenen Daten las-

sen sich kaum auf den Hochschulbereich übertragen. Aus der Hochschule steht für den deutschen Bereich nur unzureichendes oder unzureichend statistisch aufbereitetes Material zur Verfügung.

Die auch in der Hochschule feststellbaren Abweichungen der empirischen Mittelwerte vom theoretischen Mittelwert der Skala weisen nicht immer in dieselbe Richtung. In vielen Prüfungen sind Verschiebungen der Verteilungen zum positiven Pol zu beobachten (HITPASS 1967; LORENZ 1943). Durchfallquoten in Zwischenprüfungen deuten aber auch die Möglichkeit von Verteilungen an, die mehr schlechte Noten aufweisen als Normalverteilungen (DÄUMLING 1968; PIEHL 1972). Die Verteilungen der Noten sind über längere Zeiträume nicht stabil (AMELANG 1973; MÜLLER 1932); dabei sind im Fach Psychologie in den letzten Jahren deutliche Verbesserungen der Noten eingetreten (AMELANG 1973; GASCH 1971) (s.a. Teil 2).

Über die Verteilungsform sind im Hochschulbereich keine Untersuchungen angestellt worden. Ältere Statistiken wie die von MÜLLER (1932) und LORENZ (1943) weisen sehr sprunghafte Verteilungen aus, die auch bei reduzierter Zahl der Skalenstufen kaum die Voraussetzungen einer Normalverteilung erfüllen.

Die folgende Tabelle gibt die Verteilung der Noten in den einzelnen "Fach-Prüfer-Kombinationen" wieder; darüber hinaus die Verteilung "Über-alle-Fächer", also aller in den Fächern der mündlichen Prüfungen vergebenen Noten.

Tabelle: III.I. - 1

Verteilungen, Mittelwerte und Standardabweichungen der Noten der mündlichen Prüfungen pro "Fach-Prüfer-Kombination" und "Über-alle-Fächer"

	Notenstufen 1	2	3	4	5	arithm. Mittelwert	Standardabw.
Vordiplom							
Fach-Prüfer-Kombinationen:							
Method./02	8	5	2	1		1.75	0.93
Ausdr./09	6	4	6	2		2.22	1.06
Allg.Psychol./12	25	5	3			1.33	0.65
Allg.Psychol./14	6	0	1			1.29	0.76
Entw.Psychol./05	33	7	1	1		1.29	0.64
Pers.Psychol./02	9	3				1.25	0.45
Pers.Psychol./05	18	5	4	2	1	1.77	1.14
Physiol./08	12	16	7	5		2.13	0.99
Philos./01	5	5				1.50	0.53
Philos./03	8	10	3			1.76	0.70
Humang./07	14	11	3	1		1.69	0.81
Über-alle-Fächer	158	80	33	12	1	1.65	0.87
Hauptdiplom							
Fach-Prüfer-Kombinationen:							
Diagnostik/06	19	6	1	0		1.31	0.55
Sozialpsychol./05	13	2	0			1.13	0.35
Sozialpsychol./12	13	0				1.00	0.00
Päd.Psychol./02	12	1	1			1.21	0.58
Päd.Psychol./09	9	5	0			1.36	0.50
Klin.Psychol./06	8	3	0			1.27	0.47
Klin.Psychol./18	15	0				1.00	0.00
Angew.Psychol./02	4	1	0			1.20	0.45
Angew.Psychol./04	5	2	1	0		1.50	0.76
Angew.Psychol./14	13	1	1	0		1.20	0.56
Über-alle-Fächer	119	22	4			1.21	0.47

Methodisches Vorgehen: Es sind die absoluten Häufigkeiten der einzelnen Notenstufen eingetragen. Der arithmetische Mittelwert und die Standardabweichung in den letzten beiden Spalten sind wegen der ungeklärten Skalenstruktur und der Verteilungsform nur bedingt aussagekräftig.

Die Notenverteilungen in den einzelnen "Fach-Prüfer-Kombinationen" weisen in den meisten Fällen die Note 1 als häufigste Note aus. Sie ähneln in der Form Verteilungen seltener Ereignisse, den Poisson-Verteilungen. Eine statistische Überprüfung der Abweichung von Normalverteilungsmodellen erübrigt sich, wenn die Normalverteilung über die fünf Stufen der Notenskala errichtet wird.

Die besonders extreme Form der Notenverteilungen im HD ist auch auf die Wirkung der Selbstselektion der Prüflinge bei der Beantwortung zurückzuführen. Sie ist nicht für die Notenverteilung der Prüfungen dieser drei Prüfungstermine repräsentativ (s. Exkurs I).

WOTHKE (1974) stellt bei der Auswertung des vollständigen Notenmaterials der drei Prüfungstermine fest, daß der für die einzelnen Fächer berechnete Informationsgehalt der Notenverteilungen im VD zwischen 0.7 und 1.7 bit liegt, während er im HD sich zwischen 1.0 und 1.6 bewegt. (Zur Veranschaulichung: ein Informationsgehalt von 1 bit besagt, daß dieselbe Information durch zwei gleichwahrscheinliche Notenstufen zu gewährleisten wäre.)

Eine Bewertung der Ergebnisse im Hinblick auf Implikationen für mögliche Wirkungen erfolgt im Teil 4.

Bildung von Gesamtnoten

Schon BADER (1913) weist auf die Problematik der Bildung von Gesamtnoten durch arithmetische Mittelung hin, indem er die Summierung von Ungleichem

oder die zu starke Berücksichtigung von Extremen kritisiert. Noch formaler läßt sich allein aus einer statistischen Betrachtung einiges gegen die Berechnung von Gesamtnoten durch einfache arithmetische Mittelung einwenden. Die Einwände gelten auch dann, wenn die gemittelten Noten intervallskaliert sind.

1. Die Höhe der Gesamtnote wird bedingt durch den Mittelwert der Verteilungen der Einzelnoten. Das ist dann problematisch, wenn für bestimmte Prüflinge Einzelnotenverteilungen andere Mittelwerte haben als für andere. Das ist dadurch möglich, daß die Prüflinge in einem Fach von verschiedenen Prüfern geprüft werden (s. Teil 3). Eine Angleichung des Mittelwerts der Einzelnotenverteilungen kann dem Rechnung tragen (ATHANASIADIS 1974). Sie ist in der Prüfungsordnung nicht vorgesehen (und wegen der geringen Prüflingszahlen auch nicht sinnvoll).

2. Sind die Mittelwerte in den Einzelnotenverteilungen, aus denen die Gesamtnoten gebildet werden, gleich, geht trotzdem nicht jede Einzelnote mit dem gleichen Gewicht in die Gesamtnote ein. Statistisch gesehen hängt der Anteil einer Einzelnote von ihrer Streuung ab. Stärker streuende Einzelnoten bestimmen den Gesamtwert bedeutsamer als weniger streuende. Dem kann durch eine Angleichung der Streuung der Noten vor der Summenbildung Rechnung getragen werden (GLOCK & MILLMAN 1968; DAVIS 1964, der zudem für unterschiedliche Korrelationen eine Gewichtung nach Varianz oder nach Standardabweichung vorschlägt). Eine Angleichung der Streuung ist ebenfalls nicht vorgesehen.

3. Immer dann, wenn Mittelwerte aus Einzelwerten gebildet werden, die nicht vollständig miteinander korrelieren, was eine Mittelwertsbildung außerdem überflüssig machen würde, hat das Ausmaß der Korrelation zwischen den Einzelwerten einen Einfluß auf die Streuung der Mittelwerte (MEHRENS 1970;

THORNDIKE 1972). Die Streuung der Mittelwerte sinkt mit abnehmendem oder negativ werdendem Zusammenhang deutlich. Diesem Einfluß kann nur durch eine Neustandardisierung der Mittelwerte Rechnung getragen werden. (Auch sie findet nicht statt.)

Die Bildung von Notenmittelwerten als harmonischen Mitteln, die HAUSNER (1973) vorschlägt, trägt diesen Bedenken allesamt keine Rechnung und kann deshalb auch keine diskutierbare Alternative darstellen.

Über die rein formal-statistische Betrachtung hinausgehend ist außerdem zu berücksichtigen, daß die Gesamtnote durch die Struktur der Beziehungen zwischen den Einzelnoten (in der Terminologie der dafür häufig angewendeten Faktorenanalyse: Faktorenstruktur) bedingt wird. Die Kenntnis der Struktur ist notwendig, um die inhaltliche Bedeutung der Gesamtnote richtig einzuschätzen (wobei noch fraglich ist, ob überhaupt eine für alle Prüflinge gleiche Struktur angenommen werden darf).

Tabelle: III.I. - 2

Verteilungen der Gesamtnoten
(Zusätzlich: "Weite Stichprobe")

	Notenstufen 1	2	3	arithm. Mittelwert	Standardabweichung
Vordiplom					
	25	15	3	1.49	0.63
"Weite Stichprobe"	32	22	6	1.57	0.67
Hauptdiplom					
	25	4		1.14	0.35
"Weite Stichprobe"	34	12		1.26	0.44

Die Verringerung der Streuung der Gesamtnote (s. 3.) gegenüber der Streuung der Einzelnoten ("Über-alle-Fächer"; s. Tabelle III.I. - 1) wird an der Stan-

dardabweichung deutlich. Im HD streut die Gesamtnote so wenig, daß nur eine Auswertung in der unter Einbeziehung der Voruntersuchungen entstandenen "Weiten Stichprobe" möglich scheint. Die Ergebnisse werden durch die Untersuchung von WOTHKE (1974) bestätigt: Der rein quantitative Informationsgehalt der Gesamtnoten ist geringer als der der meisten Einzelnoten.

Durch die Korrelation der Einzelnoten mit den Gesamtnoten kann das Ausmaß der Berücksichtigung der Einzelnoten in den Gesamtnoten und damit zugleich der Einfluß der Streuung der Einzelnoten auf die Höhe der Gesamtnote (s. 2.) empirisch festgestellt werden. Es ist allerdings zu berücksichtigen, daß der Zusammenhang, wie dargestellt, außerdem von der dimensionalen Struktur der Einzelnoten abhängig ist.

Tabelle: III.I. - 3

Kontingenzen zwischen Noten der Einzelfächer/Hausarbeit und Gesamtnoten im VD

Fächer	Produkt-Moment-Korrelation	PHI-Korrelation	arithm. Mittelw. der Einzelnoten	Standardabweich. der Einzelnoten
Ausdr./Method.	.62	.58	1.95	0.98
Allg.Psychol.	.59	.48	1.33	0.64
Entw.Psychol.	.67	.43	1.28	0.63
Pers.Psychol.	.58	.52	1.60	1.00
Physiol.	.42	.29	2.13	0.99
Wahlfach I	.41	.28	1.68	0.67
Wahlfach II	.24	.21	1.66	0.77
Hausarbeit (nur APO)	.71	.60		

Methodisches Vorgehen: Es werden sowohl Produkt-Moment-Korrelationen als auch, wegen der Eigenart der Verteilung, parameterfreie PHI-Korrelationen durch Unterteilung der Noten in die Stufen "1" und "schlechter als 1" berechnet. Signifikanzen werden nicht angegeben,

weil die Gesamtwerte nicht bereinigt sind, d.h. der jeweilige Einzelwert, mit dem sie korreliert werden, ist in ihnen enthalten. Die Ergebnisse können nicht mehr als grobe Hinweise sein. (So wird auch nicht zwischen NPO und APO getrennt, bei denen die Bestimmung der Gesamtnoten auf unterschiedlicher Grundlage erfolgt.)

Aus der Tabelle: III.I. - 3 wird deutlich, daß der Zusammenhang von Einzelnoten und Gesamtnoten in einzelnen Fächern unterschiedlich ist. Läßt sich der geringere Zusammenhang der beiden Wahlfächer noch aus der geringeren Gewichtung erklären, ist dies für das Fach Physiologie nicht möglich. Dort ist aber auch nicht etwa die Streuung geringer, sondern größer, so daß nur die dimensionale Struktur der Einzelnoten ein Grund sein kann. Auch bei den psychologischen Einzelfächern scheint die Streuung der Fächernoten nicht allein das Ausmaß des Zusammenhangs zu determinieren.

Die entsprechende Auswertung für die Daten des HD kann nur an der "Weiten Stichprobe" erfolgen, weil die Gesamtnoten ohne Einbeziehung der Voruntersuchungen zu wenig streuen.

Tabelle: III.I. - 4

Kontingenzen zwischen Noten der Einzelfächer/Hausarbeit und Gesamtnoten im HD (nur "Weite Stichprobe")

Fächer	PHI-Korrelation	Häufigkeiten Einzelnote 1	Häufigkeiten Einzelnote >1
Diagnostik	.48	31	15
Sozialpsychol.	.62	37	8
Päd.Psychol.	.01	36	10
Klin.Psychol.	.31	39	6
Angew.Psychol.	.19	35	11
Hausarbeit (nur NPO)	.50		

Methodisches Vorgehen: Es gelten dieselben Einschränkungen wie für Tabelle: III.I. - 3.

Die Einzelnoten weisen trotz gleicher Gewichtung einen Zusammenhang unterschiedlichen Ausmaßes mit der Gesamtnote auf. Die Differenziertheit der Verteilung der Einzelnoten kommt als Grund dafür nicht in Frage; auch hier scheint die dimensionale Struktur der Fächer ausschlaggebend.

Transformation der Noten

In der vorliegenden Literatur werden drei verschiedene Verfahren der Transformation von Noten beschrieben, die teilweise miteinander kombinierbar sind. Es sind zu unterscheiden lineare Transformationen von Flächentransformationen. Die linearen Transformationen wiederum unterteilen sich in solche, die die gegebenen Verteilungen in festgesetzte Mittelwerte und/oder Streuungen umrechnen, und solche, die bei dieser Umrechnung von Schätzungen der wahren Mittelwerte und/oder Streuungen ausgehen.

Die einfache lineare Transformation, die bestimmte Einzelnotenverteilungen einander angleicht (ASCHERSLEBEN 1970; GLOCK & MILLMAN 1968), ist an das Vorliegen sehr umfangreicher Stichproben gebunden, da sie sonst zu fehlerhaften Werten führt (INGENKAMP 1975).

Demgegenüber ist die besonders in der amerikanischen Literatur empfohlene lineare Transformation mit einer Schätzung der Parameter aus externen Quellen schon auf kleinere Gruppen anwendbar. Die Mittelwerte und/oder Streuungen einer Notenverteilung werden durch Regressionsschätzungen aus den Ergebnissen von allgemeinen Leistungstests oder Intelligenztests oder anderen Leistungsdaten gewonnen (DAVIS 1964; HILLS 1972; McINTOSH et al. 1962).

Die dritte Form, die Flächentransformation, stellt ein Verfahren dar, mit dem die Noten einem Prozeß der Skalierung unterworfen werden, indem die vorliegende Häufigkeitsverteilung in eine Normalverteilung umgerechnet wird. Auch dieses Verfahren ist daran gebunden, daß umfangreichere Stichproben vorliegen. Bis auf GUTJAHR (1974) und ORLIK (1961) liegen kaum Hinweise auf die funktionale Bedeutung der Notenskalierung vor.

Insgesamt fehlen differenzierte Zuordnungen von Transformationsverfahren und Charakteristik der Notenverteilungen; Ansätze dazu machen McINTOSH et al. (1962).

Die Notenverteilungen dieser Arbeit legen wegen der geringen Zahl von Personen und des Fehlens von weiteren Leistungsdaten einen Verzicht auf eine Notentransformation nahe. Doch ist noch zu differenzieren. Notentransformationen werden gewöhnlich auf der Objektseite vorgenommen und dienen als Grundlage der Rückmeldung. Die Rückmeldung aber ist schon erfolgt. Die Anwendung einer Transformation gewinnt damit eine andere Bedeutung: Sie stellt den Versuch dar, den Verarbeitungsprozeß, in dem die Prüflinge den objektiven Noten Informationen entnehmen, nachzubilden. Für eine solche Nachbildung sind zwar die vorgestellten Modelle nicht entwickelt, es soll aber mit dem einfachsten Modell der Umrechnung auf gleiche Mittelwerte und Streuungen durch lineare Transformation der Noten der einzelnen "Fach-Prüfer-Kombinationen" der Versuch einer Nachbildung der Verarbeitung der Noten durch den Prüfling gemacht werden. Die Flächentransformation ist wegen der geringen Notenstreuung dazu weniger geeignet; für eine Schätzung der "wahren" Parameter fehlen die Vergleichsdaten.

Zusammenfassung (III.I.B.1.)

Die praktische Verwendung der Noten nimmt

eine gleichabständige Intervallskala an; in der Literatur wird den Noten nur eine Ordinalskalierung zugebilligt. "Direkte" empirische Versuche einer Bestimmung der Skalenstruktur sind selten. Sowohl die SEU-Skalierung als auch der Paarvergleich deuten auf große individuelle Unterschiede subjektiver Notenskalen der Prüfer hin. Gleichabständigkeit der Skalenstufen scheint nicht gegeben zu sein.

Die Frage der Verteilungseigenschaften wird von der Diskussion um die Skaleneigenschaften abgelöst. Ein präskriptives Normalverteilungsmodell wird abgelehnt. Die empirischen Notenverteilungen in den "Fach-Prüfer-Kombinationen" und "Über-alle-Fächer" sind extrem einseitig; fast immer ist die Note "1" die häufigste Note. Die Verteilungen ähneln Poisson-Verteilungen (Verteilungen seltener Ereignisse).

Die Bildung von Gesamtnoten ist rein statistisch gesehen problematisch: Unterschiedliche Mittelwerte der "Fach-Prüfer-Kombinationen" beeinflussen die Gesamtnote; das Ausmaß der Streuung der Einzelnotenverteilungen bestimmt ihre Berücksichtigung in der Gesamtnote; die Streuung der Gesamtnoten ist geringer als die der Einzelnoten. Darüber hinaus beeinflußt auch die inhaltliche Struktur der Beziehungen der Noten ihre Berücksichtigung in der Gesamtnote. An den empirischen Gesamtnoten können diese Effekte teilweise deutlich nachgewiesen werden.

Die unterschiedlichen Methoden zur Transformation von Noten (lineare Transformation, Angleichung an externe Parameter, Flächentransformation) sind in dieser Untersuchung nur als Modelle der subjektiven Notenverarbeitung brauchbar. Als solches kann allein die lineare Transformation der Noten der "Fach-Prüfer-Kombinationen" auf gleiche Mittelwerte und Streuungen verwendet werden, weil die Voraussetzungen für die beiden anderen Verfahren noch weniger oder gar nicht erfüllt sind.

III.I.B.2.: Konstitution der Noten in Prüfungen

Mit dem Begriff der Konstitution ist der doppelte Aspekt angesprochen, unter dem weitere funktionale Eigenschaften der Noten zu analysieren sind. Noten stellen ein Produkt dar, das mit bestimmten Eigenschaften ausgestattet ist. Nicht alle Eigenschaften jedoch lassen sich an dem Produkt selbst erkennen, sondern sind erst aus dem Prozeß zu gewinnen, in dem die Noten entstehen. Für die klassischen Kriterien von Messungen bedeutet das: Während Zuverlässigkeit eindeutig als Eigenschaft des Produkts definierbar ist, kann die Validität je nach Ansatz verschieden sowohl am Produkt als auch ausschließlich am Prozeß festgemacht werden.

Da die inhaltliche Streubreite der Arbeiten zu den Eigenschaften des Produkts (Noten) oder des Prozesses (Prüfungen) groß ist, ist eine Einschränkung unerläßlich. Sie soll in mehrerlei Hinsicht vorgenommen werden. Die Darstellung befaßt sich ausschließlich mit den Ergebnissen empirischer Analysen, die zwar bei weitem den Gegenstand nicht vollständig abdecken, aber den spezifischen Beitrag der Psychologie darstellen. Weil Untersuchungen zu bestimmten Eigenschaften von Noten und Prüfungen je nach Verschiedenheit der einbettenden Situationsbedingungen in Form unterschiedlicher Bildungssysteme, Prüfungsmodalitäten usw. zu voneinander abweichenden Ergebnissen kommen, werden bevorzugt solche Untersuchungen herausgegriffen, deren Bedingungen denen dieser Untersuchung vergleichbar sind.

In der Abstimmung des Ziels der Analyse mit dem Material ordnen sich die Ergebnisse zu drei Gruppen: Eine erste Gruppe erfaßt den Anteil der Notenvarianz, der als situationsunabhängig gelten kann (Zuverlässigkeit). Eine zweite Gruppe macht einzelne Fehlerquellen namhaft, die die Noten verfälschen können.

Eine dritte Gruppe schließlich versucht, den Inhalt der Noten zu bestimmen (Validität).

Zuverlässigkeit von Noten

In den 30er Jahren werden im Rahmen einer Konferenz, an der viele führende Psychologen Europas sich beteiligten, erhebliche Zweifel an der Zuverlässigkeit der Noten, die in den bestehenden Prüfungen vergeben werden, angemeldet (HARTOG & RHODES 1936). HARTOG & RHODES berichten von einer mündlichen Prüfung im Rahmen der Zulassung zum Civil Service, in der zwei Prüfungskommissionen eine Übereinstimmung von .41 (Produkt-Moment-Korrelation; im folgenden: r =) ihrer Urteile aufweisen. Gerade auch die mündlichen Prüfungen setzen sich einer Kritik an der fehlenden Reproduzierbarkeit der auf ihrer Grundlage gebildeten Urteile aus. BARNES & PRESSEY (1929) lassen eine mündliche Prüfung als Rollenspiel durchführen. Die Übereinstimmung zwischen den einzelnen Prüfern ist gering, die Übereinstimmung von Prüfungskomitees dagegen besser, aber immer noch unzureichend. TRIMBLE (1934) dagegen weist gleichfalls an einer Simulation einer Abschlußprüfung bei festem Frageschema eine ziemlich hohe, aus der Übereinstimmung einzelner Prüfer geschätzte Übereinstimmung von Komitees mit drei Mitgliedern (r = .59) nach.

Die Forschung in Hochschulen hat sich seit diesen frühen Untersuchungen aber bei mündlichen Prüfungen nicht so wie in anderen Bereichen (Klausuren, Tests) entwickelt, weil mündliche Prüfungen in den USA nur noch wenig Bedeutung haben und auf die Abschlußprüfungen in höheren Ausbildungsabschnitten beschränkt sind (CHARVAT et al. 1968; DRESSEL 1961). Selbst dort sind sie relativ selten (KHANA & KHANA 1972). Auch im englischen Hochschulsystem stellen mündliche Prüfungen eine Ausnahme dar (COX 1966/67b).

In Deutschland dagegen, in dessen Hochschulen sie weit verbreitet sind, fehlt(e) die Bereitschaft zur empirischen Hochschulforschung.

Ein einziger Bereich, der medizinischer Fachprüfungen in den USA, ist neuerlich mit sorgfältigen, teilweise experimentell kontrollierten Ansätzen überprüft worden. Aus ihm stammt der überwiegende Teil der folgenden Ergebnisse. Es ist anzunehmen, daß diese, auf die vorliegenden Prüfungen übertragen, eher Überschätzungen der Zuverlässigkeiten darstellen, da die in Medizinprüfungen erfaßten "Strukturbildungen" selbst dann, wenn sie die Bewältigung praktischer Aufgaben vorsehen, weniger Freiheitsgrade in der Beurteilung bieten dürften als die "Strukturbildungen" in psychologischen Fachprüfungen. Darüber hinaus sind in mehreren der Untersuchungen die Prüfungen sorgfältig geplant, was die Schulung der Prüfer einschließt.

Bei der Bestimmung der Zuverlässigkeit sind zwei verschiedene Ansätze zu unterscheiden. Der erste bestimmt die Zuverlässigkeit aus der Übereinstimmung von zwei Prüfern in derselben Prüfung (was auch als Objektivität bezeichnet wird). Die Übereinstimmungen der Urteile von zwei Prüfern ergeben Werte zwischen $r = .60$ und $r = .80$ (BULL 1956; CARTER 1962; EVANS et al. 1966; KELLEY et al. 1971). Ähnliche Werte werden erreicht, wenn ein Prüfungsprotokoll nach einiger Zeit von demselben Prüfer noch einmal benotet wird (EVANS et al. 1966). Mit diesen Untersuchungen wird aber nur die Fehlervarianz erfaßt, die auf die Urteilsbildung zurückgeht.

Ein weiterer Anteil von Fehlervarianz entsteht daraus, daß die mündliche Prüfung als Interaktion zwischen Prüfer und Prüfling situativ sehr variabel ist und daß nur ein bestimmter Ausschnitt aus Strukturbildungen im Rahmen der Prüfung zum Gegenstand der Bewertung wird. (Das letztere gilt für andere Prüfungsformen ähnlich.) Deshalb vergleicht ein zweiter

Ansatz Prüfungen derselben Prüflinge bei zwei Prüfern oder Prüfungskommissionen. Es ist selbstverständlich, daß die Situation als Fehlerquelle die Zuverlässigkeit weiter absinken läßt. So schwanken die berichteten Werte zwischen r = .30 und r = .60 (BULL 1956; CARTER 1962; EVANS et al. 1966; KELLEY et al. 1971; LEVINE & McGUIRE 1970; POKORNY & FRAZIER 1966). Die Zuverlässigkeit kann allerdings durch die Zusammenfassung von Prüfungen zu demselben Gegenstand (drei Prüfungen bei jeweils zwei Prüfern) auf Werte um r = .80 ansteigen (CARTER 1962).

Für die Noten als Rückmeldungen muß die Zuverlässigkeit aus der Übereinstimmung zwischen verschiedenen Prüfungen bestimmt werden, da der Prüfling eine für die Leistungsgüte in einem Fach in seiner Gesamtheit repräsentative Information objektiv benötigt. Wird berücksichtigt, daß die Prüfungen im Fach Psychologie nicht so sorgfältig angelegt sind, ist kaum mit einer Zuverlässigkeit einer einzelnen Prüfung über r = .40 zu rechnen.

Die anschauliche Bedeutung einer solchen Angabe ist für Notenverteilungen der vorliegenden Art kaum abzuleiten (der Wert der Zuverlässigkeit ist auch nur auf "normale" Notenverteilungen übertragen zu verstehen). Die Berechnungen von Fehlurteilen gehen von Normalverteilungen der Noten aus. Nach PLEASE (1971) ist bei einem 5-stufigen System schon bei einer Zuverlässigkeit von r = .80 mit einem Anteil von mehr als 50 % Fehlurteilen zu rechnen. Bei geringeren Zahlen von Notenstufen geht natürlich die Zahl der Fehlurteile bedeutsam zurück (THORNDIKE & HAGEN 1969), dafür sind die dann vorkommenden aber um so gravierender.

Auf jeden Fall ist anzunehmen, daß ein wesentlicher Anteil der schon nicht großen Varianz der Noten in den einzelnen mündlichen Prüfungen in Unterschieden der einzelnen Prüfer und in den Zufälligkeiten der

konkreten Prüfungssituation begründet ist.

Schätzungen für die Zuverlässigkeit von Gesamtnoten, die sich aus heterogenen Leistungen wie mündlichen Prüfungen, Klausuren und schriftlichen Arbeiten zusammensetzen, sind kaum vorhanden. Nach den Ergebnissen von CARTER (1962) und KVALE (1972) ist mit deutlich höheren Zuverlässigkeiten als denen von einzelnen mündlichen Prüfungen zu rechnen.

Dieser erste methodische Ansatz stellt aber nur fest, ob in verschiedenen Prüfungen übereinstimmende Urteile gefällt werden. Es gibt darüber hinaus noch Fehlerquellen, die in dieser Berechnung der Zuverlässigkeit nicht berücksichtigt werden.

Nachweis einzelner Fehlerquellen

Der Nachweis einzelner Fehlerquellen stellt schon einen Übergang zur Analyse des Inhalts der Noten dar, indem nicht der gesamte Fehleranteil zahlenmäßig beziffert wird, sondern einzelne Fehler namhaft gemacht werden. Im Hinblick auf die Wirkungen der Noten auf die Prüflinge ist dieser Ansatz bedeutsamer als die Bestimmung der Zuverlässigkeit: Für den Prüfling als funktionalen oder rationalen Abnehmer der Note scheidet sich diese nicht wie für den Testtheoretiker in einen wahren und einen falschen Bestandteil. Auch wenn die Note nur teilweise berücksichtigt wird, sind in dieser partiellen Wirkung immer noch die Fehleranteile untrennbar mit den wahren Anteilen verbunden. Deshalb ist eine Kenntnis über die Richtung der Fehler wesentlich, um die Richtung möglicher Verfälschungen bei den Wirkungen der Noten abschätzen zu können. Gerade hier aber ist der, auch im Hinblick auf mögliche Reformen, erwünschte Beitrag empirischer Forschung unzureichend (INGENKAMP 1975). Es fehlen

solche Untersuchungen wie die von LINN et al. (1972) für Klausuren, die die Verankerung des Urteils in bestimmten Eigenschaften der beurteilten Texte nachweist.

BETZ (1974) zeigt, daß die Urteile in mündlichen Prüfungen im Fach Psychologie über den Tageslauf hinweg nach gesetzmäßigen rhythmischen Schwankungen verlaufen. Durch harmonische Analysen kann er große Teile der Urteilsvarianz aufklären. Neben solchen eher "unbewußten" Faktoren kann aber auch das individuell planbare Verhalten des Prüfers unterschiedlich sein. Eine Befragung von universitären Prüfern durch TRAVERS & GRONLUND (1950) macht die großen Unterschiede deutlich, die in bezug auf eine Einschätzung der Bedeutung von Noten, der einzelnen Notenstufen und der sinnvollen Grundlagen von Noten bestehen können.

Umfangreicher ist die Gruppe von Untersuchungen, die solche Fehlerquellen berichten, die in den Bestimmungen der Zuverlässigkeit keine Berücksichtigung finden. Eine erste Fehlerquelle sind Abweichungen der Mittelwerte der Notenverteilungen einzelner Prüfer, die sich in Korrelationen als Maßen der Zuverlässigkeit nicht bemerkbar machen (AMELANG 1973; ATHANASIADIS 1974; KVALE 1972). Nach COX (1966/67b) erklären sie insgesamt zwischen 6 und 12,5 % der Varianz der Bewertungen in Prüfungen; bei mündlichen Prüfungen dürfte der Anteil eher höher liegen.

Eine zweite Fehlerquelle sind Mittelwertsunterschiede, die sich zwischen einzelnen Prüfungsfächern und unterschiedlichen Instituten nachweisen lassen und mit Interaktionen beider Faktoren verbunden sein können (BASS 1951; GASCH 1971).

Außer diesen querschnitthaften Unterschieden sind auch "historische" Tendenzen beobachtbar, die kaum durch entsprechende Leistungsunterschiede erklärt

werden können. Die Mittelwerte der Verteilungen fallen ab (Staatsexamen für Lehrer, CAPPEL 1967; Betriebswirte, OTTWASKA 1971) oder steigen an (Psychologie, AMELANG 1973). Auf der anderen Seite bleiben die Mittelwerte gleich, obwohl in Tests Leistungssteigerungen festzustellen sind (AIKEN 1963; COX 1966/67b; JUOLA 1968).

Inhalte von Noten

Im Versuch, das einzugrenzen, worüber Noten ein Urteil fällen, sind zwei Gruppen von Untersuchungen anzutreffen. Eine Gruppe ermittelt den Zusammenhang von Noten und zukünftigen Leistungen, also die prognostische Validität von Noten. Ihre Ergebnisse sind ausschließlich für die Inhalte von Gesamtnoten von Bedeutung.

Die prognostische Validität von Gesamtnoten in Zwischenprüfungen wird an ihrer Vorhersage der Noten in Abschlußprüfungen gemessen. Auf Grund der mäßigen Zuverlässigkeit der einzelnen Noten, der Problematik von Gesamtnoten (s. oben) und der unterschiedlichen Inhalte der beiden Prüfungen sind Zusammenhänge nur mittleren Ausmaßes zu erwarten. Die ermittelten Werte belaufen sich im Fach Psychologie (GASCH 1971) zwischen VD und HD auf r = .56; im Fach Physik (KOPFMANN 1969) zwischen VD und HD auf r = .66. Die Übereinstimmungen zwischen theoretischen und praktischen Prüfungen wie der 1. und 2. Lehrerprüfung sind geringer: CAPPEL (1967) berichtet ein r = .42; KRUMM (1969) bei Handelslehrern ein r = .43. Mögliche Studienabbrüche lassen die Zusammenhänge eher als unterschätzt erscheinen; doch ist auch zu argumentieren, daß Mechanismen wie der der sich selbst erfüllenden Prophezeiung eine Überschätzung bewirken können.

Ähnliche Bedenken liegen noch näher bei der

Ermittlung prognostischer Validität der Gesamtnoten von Abschlußprüfungen. Hinzukommt das Problem, einen sinnvollen und selbst zuverlässigen und validen Indikator für den Leistungserfolg im Beruf zu finden (KVALE 1972). Da darüber hinaus die Anforderungen im Berufsbereich anders beschaffen sind als in akademischen Prüfungen, ist kaum eine bedeutsame prognostische Validität anzunehmen. Dem entsprechen die vorliegenden Ergebnisse (GASCH 1971; OTTWASKA 1971). Allein in bezug auf die weitere Tätigkeit im akademischen Bereich sind geringfügige Zusammenhänge feststellbar (GASCH 1971); ähnlich gering sind die aus anderen Hochschulsystemen berichteten Werte (KVALE 1972; OWENS 1966). Dies gilt auch für College-Abschlußnoten in den USA (HOYT 1966).

Angaben zur prognostischen Validität einzelner Fächer des VD in bezug auf das HD (GASCH 1971) sind so wenig hoch wie wenig erhellend. Den Inhalt der Noten mündlicher Prüfungen läßt eine zweite Gruppe von Untersuchungen, die prozeßorientiert vorgehen, besser abschätzen. Sie stammen aus dem Bereich der Studien zu mündlichen Prüfungen im Medizinstudium in den USA. Die Untersuchungen von EVANS et al. (1966) und McGUIRE (1963; 1966) lassen die von den Prüfern gestellten Fragen durch Experten an Hand der BLOOM-Taxonomie auf die Prozesse hin beurteilen, die durch die Fragen angesprochen werden. Dabei stellt sich heraus, daß die Fragen in Prüfungen im wesentlichen das Erinnern von Fakten (recall) verlangen, während komplexere Leistungen selten angesprochen werden. Damit ist dann auch die Aussagekraft von Noten über solche Leistungen, die Ziel wissenschaftlicher Ausbildung sein sollten, gering.

LEVINE & McGUIRE (1970) können allerdings nachweisen, daß bei entsprechend sorgfältiger Planung auch komplexere Prozesse in mündlichen Prüfungen er-

faßt werden. Die Auswertung von Tonbandmitschnitten der vorliegenden Prüfungen (SCHLIEHE & SEEGER 1974) weist in den meisten Prüfungen eine derartige Häufung von Fragen nach, die kaum die Erwartung zuläßt, daß in den Prüfungen komplexere Leistungen ihren adäquaten Ausdruck finden.

Zusammenfassung (III.I.B.2.)

Die Konstitution der Noten bezieht sich auf Eigenschaften des Produkts (Noten) wie auch die des Prozesses (Prüfungen). Dabei beschränkt sich die Auswertung vorliegender Untersuchungen weitgehend auf empirische Arbeiten zu mündlichen Prüfungen.

Die Zuverlässigkeiten von mündlichen Prüfungen liegen, wenn sie als Übereinstimmung der Urteile von zwei Prüfern in derselben Prüfung bestimmt werden, zwischen r = .60 und r = .80 (Produkt-Moment-Korrelation). Sie sinken auf Werte von r = .30 bis r = .60, wenn sie als Übereinstimmung zwischen zwei Prüfungen bestimmt werden. Die für die Rückmeldung maßgebliche Zuverlässigkeit der einzelnen mündlichen Prüfungen dieser Untersuchung dürfte kaum größer als r = .40 sein. Die Zuverlässigkeit der Gesamtnoten ist sicher höher; Untersuchungen dazu liegen nicht vor.

Die Bestimmung einzelner Fehlerquellen ist für die Verwendung der Noten als Rückmeldung wesentlich. Es fehlen aber systematische Untersuchungen. Einzelne Untersuchungen belegen rhythmische Schwankungen der Notengebung im Tagesverlauf, unterschiedliche Mittelwerte bei Prüfern, Fächern und Hochschulen. Die Mittelwerte der Notenverteilungen sind über die Jahre hinweg Schwankungen unterworfen, die kaum auf Leistungsschwankungen zurückführbar sind.

Der Inhalt der Gesamtnoten kann als prognostische Validität bestimmt werden. Die Noten in Zwischenprüfungen weisen einen mäßig engen Zusammenhang mit

Noten in Abschlußprüfungen auf (r = .40 bis r = .60). Die Noten in Abschlußprüfungen sagen den Berufserfolg kaum vorher. Der Inhalt der Noten in den einzelnen mündlichen Prüfungen scheint mehr die Fähigkeit zur Wiedergabe von einzelnen Fakten als komplexere geistige Leistungen zu berücksichtigen.

III.I.B.3.: Konstitution der Noten dieser Untersuchung

Die Analyse der Noten dieser Untersuchung dient nicht dem Zweck, Erkenntnisse über Noten zu gewinnen; dann wäre eine Auswertung des vollständigen Notenmaterials aus den Prüfungsakten (WOTHKE 1974) angebracht. Sie soll vielmehr die konkreten Eigenschaften und Bedingungen der Noten erkunden, die in den folgenden Schritten der Auswertung die unabhängigen Variablen darstellen. Wegen der Eigenart der Notenverteilungen werden parameterfreie Verfahren bevorzugt angewendet.

Einfluß des Prüfers

Der Einfluß des Prüfers kann durch einen Vergleich der Notenverteilungen der einzelnen "Fach-Prüfer-Kombinationen" untersucht werden. Dabei werden die einzelnen Noten in den Fächern für die inferenzstatistische Prüfung als voneinander unabhängig betrachtet. Die vorhandene Abhängigkeit (sie ist, wie sich noch zeigen wird, nicht besonders hoch) wirkt sich konservativ aus.

Tabelle: III.I. - 5

Notenverteilung
in den "Fach-Prüfer-Kombinationen" des VD

Noten	Method./02	Ausdr./09	Allg.Psychol./12	Allg.Psychol./14	Entw.Psychol./05	Pers.Psychol./02	Pers.Psychol./05	Physiol./08	Philos./01	Philos./03	Humang./07
1	8	6	25	6	33	9	18	12	5	8	14
>1	8	12	8	1	9	3	12	28	5	13	15

Methodisches Vorgehen: Die Noten werden dichotomisiert in "1" und "schlechter als 1" (s. Tabelle: III.I. - 1).

Die Notenverteilungen der einzelnen "Fach-Prüfer-Kombinationen" des VD unterscheiden sich statistisch bedeutsam ($CHI^2 = 36.9$; $df = 10$; $p < .01$ (?)). Dieser Unterschied ist nicht allein darauf zurückzuführen, daß die nichtpsychologischen Prüfer gegenüber den psychologischen schlechtere Noten erteilen, auch die Unterschiede in den "Fach-Prüfer-Kombinationen" der psychologischen Fächer sind statistisch bedeutsam ($CHI^2 = 17.0$; $df = 6$; $p < .01$ (?)). (Das Fragezeichen soll andeuten, daß Werte in die Berechnung eingehen, die nicht voll voneinander unabhängig sind.)

Die Unterschiede zwischen den einzelnen "Fach-Prüfer-Kombinationen" geben ungetrennt die Einflüsse einzelner Fächer und die einzelner Prüfer wieder. Eine Trennung der beiden Einflußquellen ist, wie dargestellt, nur ansatzweise möglich. Der Einfluß des Prüfers könnte deutlich werden, wenn die Noten eines Prüfers mit den übrigen in einem Fach vergebenen Noten verglichen werden. Das ist dort möglich, wo ein Prüfer

nicht fast alle Prüfungen in einem Fach abnimmt, was für neun "Fach-Prüfer-Kombinationen" gilt. In keinem der Fälle kann mit dem Fisher-Yates-Test (SIEGEL 1956) ein statistisch signifikanter Unterschied ermittelt werden. (Auch eine entsprechende Prüfung in der"Weiten Stichprobe" bleibt ohne statistisch bedeutsames Ergebnis.) Die Bedeutung des Fachs kann nur für den Prüfer 05 untersucht werden, der in zwei verschiedenen Fächern prüft. Die beiden Notenverteilungen in den beiden Fächern unterscheiden sich nicht signifikant.

Die Auswertung der Notenverteilungen der "Fach-Prüfer-Kombinationen" des HD zeigt, daß auch hier die Benotung unterschiedlich erfolgt.

Tabelle: III.I. - 6

Notenverteilung
in den "Fach-Prüfer-Kombinationen" des HD

Noten	Diagnostik/06	Sozialpsychol./05	Sozialpsychol./12	Päd.Psychol./02	Päd.Psychol./09	Klin.Psychol./06	Klin.Psychol./18	Angew.Psychol./02	Angew.Psychol./04	Angew.Psychol./14
1	19	13	13	12	9	8	15	4	5	13
>1	7	2	0	2	5	3	0	1	3	2

Methodisches Vorgehen: Das Vorgehen entspricht dem bei der Auswertung der Notenverteilung in den "Fach-Prüfer-Kombinationen" des VD. Wegen der geringen Erwartungswerte muß die Signifikanzberechnung mit einem modifizierten CHI^2-Verfahren vorgenommen werden, das auf KULLBACK (LIENERT 1973) zurückgeht.

Die Notenverteilungen aller "Fach-Prüfer-Kombi-

nationen" sind auch im HD signifikant unterschiedlich ($CHI^2 = 17.7$; $df = 8$; $p < .05$ (?)). Der Versuch, zwischen Einflüssen des Prüfers und des Faches zu trennen, der sich der vom VD her bekannten Mittel bedient, bleibt ohne Ergebnis. Einzig im Fach Klinische Psychologie weicht der Prüfer 06 mit schlechteren Noten als die durch die anderen Prüfer erteilten ab, wenn einseitig getestet wird ($p < .05$; Fisher-Yates-Test).

Den Ergebnissen muß hinzugefügt werden, daß nicht auszuschließen ist, daß die aktive Wahl des Prüfers durch die Prüflinge an dem Entstehen der Unterschiede beteiligt sein kann (s. Exkurs II), wenngleich die Verantwortung der Prüfer und Fächer für diese Unterschiede näherliegt. Eine Entscheidung zwischen Prüfer und Fach als wesentlicher Einflußquelle ist empirisch auf Grund der vorliegenden Daten nicht zu treffen. Das umfangreichere Material von WOTHKE (1974) belegt, zumindest für das VD, daß Unterschiede zwischen den Prüfern in einzelnen Fächern ein statistisch bedeutsames Ausmaß haben. Das bestätigt die Entscheidung für die "Fach-Prüfer-Kombinationen" als Auswertungsebene.

Situative und personale Bedingungen der Einzelnoten

Schon die Unterschiede zwischen den "Fach-Prüfer-Kombinationen" sind im Sinne von Teil 2 eine die Validität der Noten einschränkende Fehlerquelle. Als solche sind auch mögliche Differenzierungen der Notenverteilungen nach situativen Bedingungen der Prüfungsart, der gewählten Prüfungsordnung und dem Zeitpunkt der Prüfung oder nach personalen Bedingungen des Alters und Geschlechts der Prüflinge so lange zu betrachten, wie nicht deren innerer Zusammenhang mit wirklichen Leistungsunterschieden aufgewiesen wird.

Die von der Prüfungsart bedingten Unterschiede der Notenverteilungen sind wegen der nachgewiesenen

Verzerrung der Notenverteilung im HD durch Selbstselektion entsprechend deutlich.

Tabelle: III.I. - 7

Noten "Über-alle-Fächer" bei VD und HD

		Noten 1	Noten >1
Vordiplom	Über-alle-Fächer	158	126
	nur psychol.Fächer	113	59
Hauptdiplom	Über-alle-Fächer	119	26

Methodisches Vorgehen: Die Noten sind wie in den letzten Tabellen dichotomisiert.

Der Vergleich der Notenverteilungen zeigt, daß im HD in den einzelnen Prüfungen bessere Noten erteilt werden. Der Unterschied bleibt bestehen, wenn im VD nur die Ergebnisse der psychologischen Fächer berücksichtigt werden ($p < .01$ (?); Fisher-Yates-Test).

Eine wesentliche Quelle dieses Unterschiedes ist sicher die durch die Selbstselektion "aufgebesserte" Notenverteilung im HD. Aber Auswertungen des vollständigen Notenmaterials der drei Prüfungstermine beweisen, daß auch dort Unterschiede in der aufgezeigten Richtung vorhanden sind. Wegen der größeren Häufigkeit erfolgt eine Auswertung mit dem Vierfelder-CHI-Quadrat-Test ($CHI^2 = 13.45$; $df = 1$; $p < .01$ (?)). Werden für das VD nur die Noten der psychologischen Fächer einbezogen, ist der Unterschied nicht mehr signifikant ($CHI^2 = 2.14$; $df = 1$; $p > .10$ (?)).

Ergebnisse der Auswertungen von Notenverteilungen gleicher Prüfungen an anderen Instituten der BRD stimmen nicht durchgehend mit diesen Befunden überein. DOUBRAWA (1971) nimmt eine solche Auswertung an seinen Daten nicht vor; doch läßt sich aus den wiedergegebenen Tabellen nur eine geringfügige Verschie-

bung in positiver Richtung im HD ablesen. Bei GASCH (1971) ist nur an einem von drei Instituten ein niedrigerer Notendurchschnitt im VD festzustellen.

Eine Bedingung der Notenverteilung durch die Prüfungsordnung ist nicht zu belegen.

Der Einfluß des Zeitpunkts der Prüfung deutet sich bei einem Vergleich der Noten der drei Prüfungstermine mit den Noten der Prüfungstermine der Voruntersuchung an: Die Noten werden (nicht signifikant) besser. WOTHKE (1974) kann solche Unterschiede für sein Notenmaterial nachweisen, das neben den Noten dieser drei Prüfungstermine auch noch Noten aus älteren Prüfungsakten einbezieht. Es kann dabei statistisch abgesichert werden, daß diese Verschiebungen nicht nur durch den Wechsel der Prüfer entstehen.

Die Versuche, personale Bedingungen der Einzelnoten zu überprüfen, verschärfen das inferenzstatistische Problem der Verwendung abhängiger Daten. Die Einbeziehung personaler Eigenschaften bedeutet eine vollständige Abhängigkeit auf seiten der Bedingungen für die jeweiligen Einzelnoten einer Person. Da darüber hinaus noch die Bedingungen selbst nicht voneinander unabhängig sind, sollen die Ergebnisse weder inferenzstatistisch bearbeitet noch im einzelnen interpretiert werden, da sie nicht einmal mit den Ergebnissen anderer Untersuchungen voll übereinstimmen (DOUBRAWA 1971; GASCH 1971).

Tabelle: III.I. - 8

Alter der Prüflinge und Noten in mündlichen Prüfungen

Vordiplom	Noten		Hauptdiplom	Noten	
Alter	1	≥1	Alter	1	≥1
unter 25	88	85	unter 27	67	8
gleich/über 25	70	41	gleich/über 27	52	18

Methodisches Vorgehen: Die Noten sind wie üblich, das Alter ist am Median der Altersverteilung dichotomisiert.

Tabelle: III.I. - 9

Geschlecht der Prüflinge und Noten in mündlichen Prüfungen

Vordiplom	Noten 1	Noten >1	Hauptdiplom	Noten 1	Noten >1
männlich	103	64	männlich	67	13
weiblich	55	62	weiblich	52	13

Im VD erzielen die älteren Prüflinge die relativ besseren Noten. Im HD ist der Unterschied eher geringer; dort sind es aber die jüngeren Prüflinge, die die besseren Noten aufweisen. Zwischen den beiden Geschlechtern ist nur im VD ein Unterschied festzustellen: Dort erhalten die männlichen Prüflinge die besseren Noten.

Situative und personale Bedingungen der Gesamtnoten

Zwar weisen die bei den Gesamtnoten feststellbaren Unterschiede jeweils in die von den Einzelnoten vorgezeigte Richtung, was die Form der Gesamtnotenbildung nicht anders erwarten läßt, doch ist nur eine einzige Bedingung von nachweisbar signifikantem Einfluß: die Prüfungsart, für die beim HD bessere Gesamtnoten festzustellen sind ($p<.01$; Fisher-Yates-Test). Fallen die Einflüsse der Selbstselektion weg, dann ist der Unterschied nur noch unbedeutend.

Zusammenhänge zwischen den Einzelnoten

Die Zusammenhänge zwischen den Einzelnoten wurden bereits in Teil 1 im Rahmen der Gesamtnotenbildung angesprochen. Dort wurde deutlich, daß sie wesentlich

das Ausmaß bestimmen, mit dem eine Fachnote in der Gesamtnote Berücksichtigung findet. Auch im Hinblick auf die Problematik der inferenzstatistischen Behandlung der Daten auf der Ebene "Über-alle-Fächer" ist eine Analyse der Abhängigkeiten zwischen den Noten bedeutsam. Je geringer diese Abhängigkeiten ausfallen, um so weniger problematisch ist die entsprechende Anwendung inferenzstatistischer Verfahren.

Tabelle: III.I. - 10

Zusammenhänge der Einzelfachnoten im VD und HD

Vordiplom

	Ausdr. Method.	Allg. Psychol.	Entw. Psychol.	Pers. Psychol.	Physiol.	Wahlfach I
Allg. Psychol.	.30^{+}					
Entw. Psychol.	.26	.33^{+}				
Pers. Psychol.	.28	.46^{++}	.52^{++}			
Physiol.	.26	.29^{+}	.12	.19		
Wahlf. I	.34^{+}	.18	.08	.01	.05	
Wahlf. II	.03	.19	.37^{+}	.16	.02	.21

Hauptdiplom

	Diagnostik	Sozial-psychol.	Päd. Psychol.	Klin. Psychol.
Sozialpsychol.	.33			
Päd.Psychol.	.03	.00		
Klin.Psychol.	.05	.14	.06	
Angew.Psychol.	.03	.33	.03	.21

Methodisches Vorgehen: Die Zusammenhänge werden als PHI-Koeffizienten der wie üblich dichotomisierten Noten berechnet. Die Signifikanzgrenze wird gemäß LIENERT (1973) über den Fisher-Yates-Test geprüft. Als

zweiseitige Signifikanzschranken sind Korrelationen, die mit $p < .05$ von PHI = .00 abweichen, mit "+" gekennzeichnet, solche mit $p < .01$ mit "++".

Die Korrelationen im VD sind nur mäßig hoch; die von GASCH (1971) berichteten Werte werden nicht erreicht (dort werden allerdings auch Produkt-Moment-Korrelationen berechnet). Das Muster der Korrelationen deutet eine Binnenstruktur an, die die unterschiedlichen Zusammenhänge von Einzelfachnoten und Gesamtnoten erklärbar macht. Für die Auswertung auf der Ebene "Über-alle-Fächer" sind die nicht besonders hohen Korrelationen ein Hinweis darauf, daß die Verfälschung der inferenzstatistischen Verfahren durch die partielle Abhängigkeit der Noten nicht sehr gravierend ausfallen dürfte.

Die Korrelationen im HD sind nur gering; die von GASCH (1971) errechneten Zusammenhänge sind deutlich enger (sie streuen zwischen r = .20 und r = .55). Da im HD die Abweichungen von der Notenstufe "1" auch bei den Einzelfachnoten selten sind, wäre eine bedeutsame Abhängigkeit ein Einwand gegen die Auswertung "Über-alle-Fächer", da dann nur wenige Personen über die Abweichungen mit der Varianz auch die Ergebnisse determinieren könnten. Die folgende Aufstellung soll verdeutlichen, daß eine größere Zahl von Prüflingen an den Abweichungen beteiligt ist.

Tabelle III.I. - 11

Einzelfachnoten der Prüflinge im HD

Zahl der Einzelfachnoten >1	0	1	2	3	4
Zahl der Prüflinge	13	9	5	1	1

Methodisches Vorgehen: Es ist die Anzahl der Prüflinge dargestellt, die keine, ein, zwei, drei und vier Fächer aufweisen, in denen sie jeweils eine

schlechtere Note als eine "1" erhalten haben.

Von den 29 Prüflingen haben immerhin 16 in wenigstens einem Fach eine schlechtere Note als 1 erhalten und sind entsprechend an der Determination der Varianz (und Kovarianz) beteiligt.

Zusammenfassung (III.I.B.3.)

Es sollen die konkreten Eigenschaften der Noten ermittelt werden, die in der Untersuchung verwendet werden.

Die Notenverteilungen der einzelnen "Fach-Prüfer-Kombinationen" unterscheiden sich sowohl im VD als auch im HD statistisch bedeutsam. Es kann nicht nachgewiesen werden, ob das allein auf Unterschiede zwischen den Prüfern oder auch noch auf Unterschiede zwischen den Fächern zurückgeht.

Die Auswertung der Einzelnoten "Über-alle-Fächer" ergibt statistisch bedeutsame Unterschiede zwischen VD und HD, die selbst noch bei einer Kontrolle der Selbstselektion erhalten bleiben. Bei Personenunterschieden erfolgt keine inferenzstatistische Auswertung: Im VD erzielen die älteren Prüflinge die relativ besseren Noten; im HD sind es die jüngeren. Weibliche Prüflinge schneiden im VD schlechter ab als die männlichen, während im HD wenig Unterschied besteht.

Die Auswertung der Gesamtnoten ergibt bis auf die Unterschiede zwischen VD und HD keine weiteren Unterschiede zwischen den Bedingungen.

Die Korrelationen zwischen den Einzelnoten sind im VD nur mäßig hoch und streuen stark. Die Auswertung "Über-alle-Fächer" ist von daher gesehen nicht so problematisch. Die entsprechenden Korrelationen im HD sind durchweg gering. Eine Auswertung der von der Notenstufe "1" abweichenden Noten zeigt, daß "Über-alle-

Fächer" daran 16 der Prüflinge beteiligt sind, so daß auch hier eine inferenzstatistische Auswertung unter der Annahme der Unabhängigkeit nicht ganz so problematisch scheint.

III.I.B.4.: Eignung der Noten zur Rückmeldung

Der mögliche Sinn der Wirkung der Noten (nicht unbedingt ihre faktische Wirkung) ist durch die Information begrenzt, die die Noten bereitstellen. In Kapitel II.4. wurde der grundlegende Charakter der Information als die vergleichsorientierte, wertende Kennzeichnung der Leistungsgüte bestimmt. Die Aufarbeitung von fremden Untersuchungen und die Auswertung der Noten im Rahmen dieser Untersuchung haben deutlich werden lassen, daß selbst diese im Anspruch gegenüber dem pädagogisch Sinnvollen schon zurückgeschraubte Information durch Noten und Prüfungen von fragwürdiger Qualität ist.

Ein größerer Teil der in Noten aus mündlichen Prüfungen enthaltenen Information - vermutlich der überwiegende - geht auf die besonderen Eigenarten der Interaktionssituation bestimmter Prüfer mit bestimmten Prüflingen zurück, sowie auf die individuell unterschiedlichen Notenverteilungen der einzelnen Prüfer und offensichtlich eine Vielzahl weiterer Quellen, die eine "Fehlinformation" beinhalten. Zudem erscheint fraglich, ob die in Prüfungen dieser Art gezeigten Leistungen den Anspruch erheben können, für wissenschaftlich gerechtfertigte Strukturbildungen repräsentativ zu sein. Die eher negative Bilanz verbessert sich kaum dadurch, daß manche Fehlerquellen, wie z.B. die unterschiedlichen Notenverteilungen bei einzelnen Prüfern und einzelnen Fächern, bis zu einem bestimmten Maß bei der Umsetzung zu Selbstbewertung und Motivation von den Prüflingen ausgeglichen werden könnten.

Die aus den einzelnen Noten gebildeten Gesamtwerte ziehen dieselben Bedenken auf sich, weil die Noten aus mündlichen Prüfungen einen wesentlichen Anteil an ihnen ausmachen. Hinzu treten u.a. die folgenden Probleme: Die Gesamtnote hängt davon ab, welchen Prüfer ein Prüfling für die Prüfung gewählt hat, oder welcher Prüfer ihm zugeteilt wurde; in sie gehen die einzelnen Fächer in unterschiedlicher Gewichtung nach den statistischen Eigentümlichkeiten der Notenverteilungen und der Inhaltsstruktur der Beziehungen zwischen den Fächern ein.

Die Einwände gegen Noten lassen insgesamt gesehen eine geringe Wirkung auf Selbstbewertung und Motivation beinahe eher als erstrebenswert erscheinen als eine bedeutsame. Wenngleich eine derartige Wertung auch noch die Alternativen zu berücksichtigen hätte, die an Stelle von Noten auf Selbstbewertung und Motivation Einfluß nehmen können. Kapitel IV muß diese Diskussion wieder aufnehmen.

Vor der Analyse der abhängigen Variablen aber bedarf die besondere Eigenart der empirischen Notenverteilungen noch einer Erörterung. Die Konzentration der erteilten Noten auf die Notenstufe "1" (im HD durch die Auswirkungen der Selbstselektion zusätzlich verstärkt) könnte als eine Benotung nach Minimalkompetenz interpretiert werden. Ist eine bestimmte Kompetenz erreicht, wird jede Leistung, sei sie noch so unterschiedlich, mit der Note "1" bewertet. Eine solche Benotung nach Minimalkompetenz könnte Ergebnis einer "absoluten" Benotung im Gegensatz zu einer relativen, verteilungsorientierten sein. Damit könnten die Noten dann aber möglicherweise mehr Rationalität beinhalten, als es in Kapitel II.4. in der Festlegung der durch Noten verfügbaren Information behauptet wurde. Eine kurze Besinnung auf die Unterscheidung von kriteriumsorientierter ("absoluter") und normorientierter

(vergleichender) Messung, die in der Theorie der Schulleistungstests gemacht wird (POPHAM & HUSEK 1969; SCHREINER 1972; STRITTMATTER 1973), wirkt klärend. Das kriteriumsorientierte Konzept versucht, bei der Leistungsmessung pädagogischen Zielvorstellungen einen Vorrang einzuräumen. Eine Messung ist nur dann als kriteriumsorientiert zu betrachten, wenn sie einen expliziten Bezug zu Zielvorstellungen aufweist. Dies ist aber bei den Noten in Prüfungen, auf jeden Fall bei diesen Noten in diesen Prüfungen nicht der Fall. Dort wird die relative Leistungskennzeichnung nicht durch explizite Zielformulierung (BROWN 1971) abgelöst, sondern der Prüfer stellt mit seinen konkreten Anforderungen und Beurteilungssystemen eine implizite Zieloperationalisierung dar, die dem Prüfling aber kaum einsichtig sein dürfte. Sie ist im Hinblick auf die Verarbeitung der in den Noten enthaltenen Information unbrauchbar. Die feststellbaren Unterschiede zwischen den "Fach-Prüfer-Kombinationen" machen zudem eine konkret verteilungsorientierte Informationsentnahme (LIEN 1967; STANLEY & HOPKINS 1972) praktisch auch empfehlenswert.

Zwei weitere Aspekte der Notenverteilungen betreffen die geplante Datenauswertung.

1. Die Noten als unabhängige Variablen enthalten in dieser Art der empirischen Verteilung nur in einem beschränkten Bereich der Verteilungen Informationen (als Abweichungen von der Notenstufe "1"). Gegenüber den annähernd normalverteilten (s. unten) abhängigen Variablen ist damit eine optimale Form der Informationsübermittlung, die ähnliche Verteilungsformen voraussetzt, von vornherein ausgeschlossen.

2. Die Formulierung von Normwerten für die Wirkung auf die Motivation, die eine zweiseitige Wirkung durch zusätzliche Kompensation vorsah (s. Kapitel II.4.), ist der empirischen Verteilung der Noten nicht angemes-

sen. Da nicht jede Abweichung von der Notenstufe "1" als Grundlage kompensatorischer Motivierung behauptet werden kann, soll in der Auswertung die Einführung der Kompensation zurückgenommen werden und nur auf positive lineare Zusammenhänge hin: schlechtere Noten - weniger Motivation ausgewertet werden. (Im übrigen hätte auch die Auffassung, jede Abweichung von der Notenstufe "1" müsse kompensatorisch motivieren, über eine Prüfung des linearen Zusammenhangs zu erfolgen, der dann nur entgegengesetzt: schlechtere Noten - mehr Motivation ausfallen müßte.)

III,I,C,: Selbstbewertung: Theorie und Methode

Da gerade für die normativ zentrale Selbstbewertung in der deutschsprachigen Literatur sich kaum genügende systematische Darstellungen finden (Artikel in den entsprechenden Handbüchern von LERSCH & THOMAE (1960), BERGIUS (1964), THOMAE (1965) und GRAUMANN (1972) geben allenfalls am Rande Auskunft; als neuere Monographie wäre allein die von NEUBAUER (1976) zu nennen), muß die Darstellung der Selbstbewertung weiter ausholen. Das hat auch seinen immanenten Reiz, weil auf diesem Gebiet sich theoretisches Durcheinander und eine Vielfalt empirischer Befunde paaren.

Vielfalt von Selbst-Begriffen

Setzt die begriffliche Analyse unterhalb des Begriffes "Selbst" (GERGEN 1971), des Ich-Begriffes, des Bewußtseinsproblems und des Identitätsproblems (NEUBAUER 1976) an, trifft sie auf eine Ansammlung von Begriffsverbindungen des "Selbst" mit ganz unterschiedlichen Gegenständen, Prozessen und Tätigkeiten. Alle diese Begriffe, denen meist nur rudimentär theoretische Begründungen zugeordnet sind, haben gemein, daß ein Individuum (und eine Sozialität) über

solche kognitiven Konzepte verfügt, die es selbst als Gegenstand haben (KAMINSKI 1964). Weitgehende Übereinstimmung besteht auch darin, daß diese Konzepte aus Mitteilungen des Individuums hervorgehen oder zumindest aus ihnen erschlossen werden können.

Eine weitergehende phänomenologische Verständigung über die dabei stattfindenden Prozesse ist nicht beabsichtigt. Statt dessen wird ein "aufsteigendes" Verfahren bevorzugt, indem vorhandene begriffliche Bestimmungen gesammelt werden, um die "Angebotsseite" zu verdeutlichen, die dann mit den Anforderungen von der Fragestellung her zu konfrontieren ist. Phänomenologische Klärungen geraten mit dem Ausmaß ihrer Differenzierung in immer größere Unvereinbarkeit mit operationistischen Begriffs- und Methodentraditionen, deren Erfahrungen darin nicht mehr integrierbar sind. Ein rein operationistisch orientiertes Vorgehen allerdings muß die Gefahr sehen, die Fülle der Gegenstandsaspekte zu verfehlen. (Ein Beispiel solcher Verfehlung ist der Ansatz von DUVAL & WICKLUND 1972.)

Die Suche nach der Verwendung von Selbst-Begriffen findet sie in der amerikanischen Forschung in großer Fülle vor. Die folgende Bestandsaufnahme ist nicht vollständig; die Zuordnung der Autoren zeigt für diesen Begriffsgebrauch kennzeichnende Arbeiten an. Die Begriffe lassen sich in zwei Gruppen aufteilen, die aber nicht vollständig diskret sind.

Mehr den Inhalt sprechen die folgenden Begriffe an: self-concept (LOWE 1961; WYLIE 1961; 1968), self-regard (WYLIE 1968), self-acceptance (CROWNE & STEPHENS 1961), self-approval (RAIMY 1948), self-enhancement (BENJAMINS 1950).

Mehr dem Prozeß, in dem ein Inhalt entsteht, gilt die zweite Gruppe: self-rating (RICHARDS 1966), self-perception (RAIMY 1948), self-report (WYLIE

1961), self-estimate (ARSENIAN 1942), self-description (VELDMAN & PARKER 1970), self-awareness (BILBY et al. 1972), self-evaluation (RADLOFF 1968).

Damit ist der Verwirrung aber noch nicht genug. Die aufgezählten Begriffe werden teilweise innerhalb derselben Arbeit ohne Begründung ausgetauscht; verschiedene Begriffe werden für dieselbe oder weitgehend ähnliche Operationalisierungen verwendet; der gleiche Begriff wird sehr unterschiedlichen Operationalisierungen zugeordnet.

Ein weiteres Problem entsteht durch die Notwendigkeit sprachlicher Übertragung. Dies soll an einem Begriff verdeutlicht werden, der in die obige Liste nicht aufgenommen wurde: self-esteem. Der Begriff, der sogar als eigenständige Theorie figuriert (JONES 1973), ist nach DUIJKER &`RIJSWIJK (1975) mit "Selbstachtung" zu übersetzen. Das mag den lexikalischen Begriffsgehalt durchaus treffen, die gemeinte Bedeutung aber wäre angemessen durch den Begriff des Selbstwertgefühls (LERSCH 1962; NEUBAUER 1976) wiederzugeben. Doch bringt das eine phänomenologische Erweiterung mit sich, die wiederum über den operationistischen Gebrauch zu weit hinausgeht.

Der Grund der Begriffsvielfalt ist u.a. in der Heterogenität der theoretischen Tradition zu suchen, aus der die unterschiedlichen Selbst-Begriffe stammen (BILBY et al. 1972; GERGEN 1971). Neben den phänomenologisch-philosophischen Quellen vor der Eigenständigkeit der Psychologie als Wissenschaft sind als psychologische Richtungen so heterogene wie Gestalttheorie, Psychoanalyse und "Aufgeklärter Behaviorismus" (WYLIE 1968) zu nennen. Auch soziologische Theorien in der Folge von COOLEY und MEAD im Rahmen des symbolischen Interaktionismus und der Rollentheorie (BILBY et al. 1972) sind daran beteiligt. Mit zur Begriffsvielfalt durch Verhinderung einer eigenständi-

gen Tradition hat ferner die Tatsache beigetragen, daß der strenge Behaviorismus in der amerikanischen Psychologie eine Zeitlang derartige Begriffe ausschloß. Aber auch die deutliche Zunahme der Beschäftigung mit Selbst-Begriffen und entsprechenden Prozessen (NEUBAUER 1976), die schon oberflächlich aus der zunehmenden Berücksichtigung im Index der Psychological Abstracts erschlossen werden kann, hat nicht zu einer Vereinheitlichung geführt. Ansätze der Soziologie und Sozialpsychologie (BILBY et al. 1972; RADLOFF 1968) stehen neben solchen der Persönlichkeitspsychologie (CROWNE & MARLOWE 1964), zu denen noch Differenzierungen innerhalb dieser Bereiche hinzutreten.

Zur weiteren Aufsplitterung trägt die Verwendung in Einzelfeldern praxisbezogener Psychologie bei: der Therapie (CROWNE & STEPHENS 1961) oder der Trainingsbewegung (SHERWOOD 1967).

In der Zunahme der Verwendung von Selbst-Begriffen deutet sich aber noch über die kognitive Wende der Psychologie hinausreichend eine einheitliche Absicht an, einen Beitrag zu rationalem Problemlösen und Entscheiden (WYLIE 1968) zu leisten und zu mehr Selbstverstehen (GERGEN 1971) beizutragen; eine Absicht, die sich mit der dieser Untersuchung in Übereinstimmung befindet. Deshalb darf die Folgerung aus der begrifflichen Unklarheit nicht der Verzicht auf Selbst-Begriffe sein. Die praktische Absicht ist Grund genug, die Begriffsdifferenzierungen zu übergehen, um dadurch die unter verschiedenen Begriffen angesammelten Erfahrungen zu gewinnen. Als Basis scheinen dafür wie für die eigene empirische Untersuchung zwei Begriffe ausreichend: der die Struktur bezeichnende des Selbstkonzepts (self-concept) und der den Prozeß bezeichnende der Selbstbewertung (self-evaluation).

Selbstkonzept als Grundbegriff

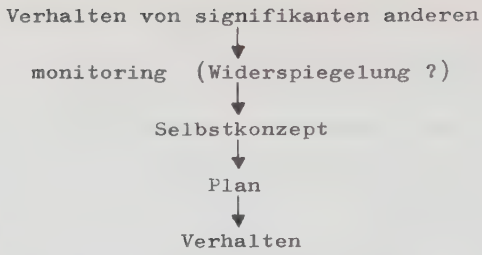
Der Begriff des Selbstkonzepts kann eine ihm zugedachte integrative Funktion nur erfüllen, wenn er sich von vereinseitigenden Operationalisierungen ablöst. Dabei kann er teilweise an seinem Ursprung anknüpfen. Der Begriff hat seinen Ursprung in der Psychologie von ROGERS. Sie gibt ihm eine, wenn auch unklare, so doch umfassende theoretische Bedeutung. In der Arbeit von RAIMY (1948; s. auch BENJAMINS 1950) wird das Selbstkonzept als organisiertes Objekt der Wahrnehmung dargestellt. Ihm kommt in der Verarbeitung vergangener und gegenwärtiger Erfahrung große Bedeutung zu, die sich auch auf das zukünftige Verhalten bezieht. RAIMY, der den Nachweis von Therapieerfolg erbringen will, arbeitet den self-approval mit Bezug auf McDOUGALL (self-regard) als eine wesentliche Dimension des Selbstkonzepts heraus. Daraus entwickelt sich mit der Integration in die akademische Persönlichkeitspsychologie eine Tradition. Mit unterschiedlichen Methoden wie Fragebögen, Eigenschaftslisten und Q-sorts (WYLIE 1968) wird die Diskrepanzdimension des Selbstkonzepts operationalisiert, indem die Differenzen zwischen realem und idealem Selbst berechnet werden. Ergebnis dieser Entwicklung ist eine Mischung aus abstraktem Konstrukt und rigider Operationalisierung in der Form solcher Diskrepanzwerte (CROWNE & STEPHENS 1961). Den Grund für die Vereinseitigung des an sich umfassenderen Begriffs vermuten die Autoren in ideologischen Quellen des Wertsystems der amerikanischen Gesellschaft.

Zwei weitere Arbeiten zur gleichen Zeit (LOWE 1961; WYLIE 1961) kritisieren begriffliche Unklarheiten und vor allem methodische Schwächen. Aber der Begriff selbst scheint ihnen einen nicht ersetzbaren Stellenwert zu haben. Eine Betonung des Strukturaspekts soll den Stellenwert konkretisieren. (WYLIE

(1968) dagegen versucht an der Lerntheorie orientiert eine prozessuale Ausarbeitung.)

Mit LOWE (1961) kann man dabei auf den Urheber (RAIMY 1948) zurückgreifen, indem das Selbstkonzept als strukturiertes Objekt der Wahrnehmung bestimmt wird, also ein kognitives Konzept darstellt. DUVAL & WICKLUND (1972) haben eine große Zahl von empirischen Untersuchungen zur Aktualgenese solcher Konzepte durchgeführt. Bei aller Skepsis gegenüber der gewählten Operationalisierung wird aus den Untersuchungen deutlich, daß die selbstbezogenen Konzepte von den auf "äußere Gegenstände" bezogenen Konzepten sich aktualgenetisch qualitativ unterscheiden.

Die Ausdifferenzierung dieses psychologischen Konstrukts (CROWNE & STEPHENS 1961; GERGEN 1971) ist durch eine Hierarchisierung formal unterschiedlicher Selbstkonzepte möglich (WYLIE 1968); im Rahmen der hier angestrebten Verwendung ist die Annahme von inhaltlichen Differenzierungen fruchtbarer. Das Selbstkonzept ist in sich organisiert (RAIMY 1948) und situationsspezifisch strukturiert; es gliedert sich in inhaltlich unterschiedliche einzelne Konzepte. (GERGEN (1971) möchte in diesem Sinne den Begriff "Konzept" eher in der Mehrzahl verwendet sehen.) Eines dieser Konzepte kann das Selbstkonzept eigener Leistungsfähigkeit darstellen. BILBY et al. (1972), die gleichfalls die Strukturierung des Selbstkonzepts betonen und dementsprechend ein solches Selbstkonzept der Leistungsfähigkeit postulieren, sehen die Spezifität von einem rollentheoretischen Ansatz her. Sie verbinden das mit dem Versuch, das Selbstkonzept der Leistungsfähigkeit in ein Ablaufmodell einzuarbeiten, das die funktionale Bedeutung des Selbstkonzepts herausstellt:



Auf der Ausgangsseite weist dieses Schema eine bestechende Nähe zur Konzeption von MILLER et al. (1973) aus, obwohl deren metaphorische Verwendung der einzelnen Bestimmungsstücke "Bild" und "Plan" einen Vergleich erschwert.

Das so bestimmte Selbstkonzept eigener Leistungsfähigkeit soll als psychologisches Konstrukt den mehr alltagssprachlich formulierten Begriff der Selbstbewertung in der folgenden Darstellung ablösen. Dazu ist eine weitere Ausformulierung auf den Ebenen Struktur (Inhalt), Prozeß und Methode notwendig.

Struktur und Prozeß beim Selbstkonzept eigener Leistungsfähigkeit

Die Ausformulierung ist auf der Strukturebene als Bestimmung des Inhalts des spezifischen Selbstkonzepts möglich und auf der Prozeßebene als Bestimmung der Beziehung, in der dieses Selbstkonzept konstituiert wird. (Diese Beziehung ist nicht sozialer Art, sondern intrapersonaler Art.) Dabei sind Struktur und Prozeß immer aneinander gebunden (GERGEN 1971): Strukturen bilden sich durch Prozesse, als bestehende bedingen sie diese wiederum. Komplizierend tritt hinzu, daß die Methode der Erfassung selbst einen Prozeß darstellt, der als empirischer mit dem Prozeß auf der Konstruktebene und der entsprechenden Struktur konfrontiert werden muß. Auf diese Weise entstehen zwei

Bezüge, in denen die Konstrukte zu sehen sind. Es handelt sich um einen empirisch-methodischen Bezug, der die Operationalisierung anspricht, und um einen begrifflich-normativen Bezug, der auf Konstruktebene zu erörtern ist. Dieser Abschnitt beschränkt sich auf den letzteren.

Die Struktur des Selbstkonzepts eigener Leistungsfähigkeit wird charakterisiert durch seine konkreten Inhalte. Die Inhalte sind durch Kapitel II. vorgegeben als wissenschaftliche Strukturbildungen im kognitiven Bereich. Deren Repräsentation auf Selbstkonzeptebene ist in zwei Richtungen denkbar. Zum einen können die Strukturbildungen in ihrer individuellen Inhaltlichkeit repräsentiert sein. Das würde bedeuten, daß sie in der Form von Theorien und Schulen der Psychologie, praktischen Tätigkeitsfeldern usw. ausformuliert werden müßten. Je weiter diese Ausformulierung einzelne spezifische Inhalte aufnimmt, um so mehr lassen sich nur qualitative Vergleiche zwischen den Selbstkonzepten der einzelnen Personen anstellen. Die Prüfung und die Noten als Mittel der Rückmeldung legen eine geringere Differenzierung nahe: die einzelnen Fächer der Prüfungsordnung (s. Teil A), die direkte Vergleiche zwischen den Prüflingen erlauben.

Ein weiterer möglicher Inhalt allgemeinerer Art ist das auf die Psychologie als Gesamtgebiet bezogene Selbstkonzept. Dabei kann noch zwischen einem gegenwartsbezogenen Selbstkonzept der augenblicklichen Leistungsfähigkeit im Gebiet der Psychologie und einem zukunftsbezogenen Selbstkonzept differenziert werden.

Die spezifischen Selbstkonzepte eigener Fähigkeiten in den einzelnen Prüfungsfächern und das gegenwarts-/zukunftsbezogene Selbstkonzept in der Psychologie insgesamt sind dann quantitativ bestimmbar, wenn sie in Eigenschaftsdimensionen fixiert wer-

den. Auch hier hat Kapitel II. schon eine Vorentscheidung getroffen: Als Dimension wurde die Güte der Leistung festgelegt.

Für die Ebene der Prozesse sind damit schon wesentliche Vorentscheidungen gefallen. Nach KAMINSKI (1971) sind verschiedene Modi der Mitteilung von Konzepten möglich. Es werden unterschieden Beschreibungen, Beurteilungen und Bewertungen. Die Modi sind auf die Konstruktebene übertragbar (zum Verhältnis von Prozeß und Mitteilung s. unten).

Die Inhalte der Selbstkonzepte, so wie sie bisher festgelegt wurden, lassen den Modus der Beschreibung nicht zu. Eine Entscheidung muß also zwischen Bewertung und Beurteilung fallen. Den normativen Vorgaben entspricht eher die Beurteilung, da sie als der "rationalere" Prozeß gelten kann. Die Diskussion wird dadurch erschwert, daß die Begriffe Beurteilung und Bewertung sehr unterschiedlich und teilweise miteinander austauschbar verwendet werden. KAMINSKI (1971) zufolge müßte die Erfassung des Selbstkonzepts eigener Fähigkeit in der Form der Beurteilung vorgenommen werden können. Für ISRAEL (1956) ist die Globalität der Erfassung aber schon der Anlaß, sie als Bewertung zu bezeichnen. Im Sinne dieses Autors stellt der angestrebte Prozeß eine informative Bewertung dar, die er von einer bestätigenden, einer wunschbezogenen und einer normativen unterscheidet. Die Annahme ist dabei aber nicht nur in der Globalität des Urteils begründet, sondern auch darin, daß die Dimension der Güte von Leistungen eine eindeutige Richtung aufweist. In bezug auf die Leistung ist das Mehr zugleich auch immer das Bessere, wie in Kapitel II. angedeutet wurde. Diese Einbettung in ein Wertsystem ist als gesellschaftlich bedingte nicht auf der Ebene der Hochschule aufzulösen. Deshalb muß der Prozeß, der das Selbstkonzept eigener Leistungsfähigkeit konstituiert, als

Bewertung bestimmt werden, wobei eine Einschränkung angebracht ist: Die Bewertung soll sich möglichst eng an Beurteilungen anschließen.

Dem kann durch die Vorgabe von expliziten Bewertungs-(Beurteilungs-)Standards Rechnung getragen werden. Standards für die Bewertung eigener Fähigkeiten sind in Form objektiver Anforderungen möglich. Liegen solche Anforderungen nicht vor, dann erfolgen entsprechende Bewertungen im Prozeß des sozialen Vergleichs (DUVAL & WICKLUND 1972; ISRAEL 1956). Da absolute Anforderungen für Fähigkeiten in bestimmten Fächern oder der Psychologie insgesamt fehlen, ist der Prozeß der Bewertung auf der Ebene des sozialen Vergleichs vorzunehmen. Um dabei möglichst konkrete Beurteilungen anzusprechen, wird auf methodischer Ebene eine Bewertung durch Einordnung in die konkrete Bezugsgruppe der Mitprüflinge verlangt. Nur beim zukunftsbezogenen Selbstkonzept eigener Fähigkeit muß eine abstraktere Bezugsgruppe vorgegeben werden.

Zur Stützung dieser Deutung des Prozesses kann die Theorie der sozialen Bewertung, die von FESTINGER (1954) formuliert wurde, nur bedingt herangezogen werden. Die Theorie ist bisher im Gegensatz zur vom gleichen Autor stammenden Dissonanztheorie empirisch eher unterbestimmt (RADLOFF 1968). Dem Versuch einer solchen bedürfnistheoretischen Grundlegung kann zudem entgegengehalten werden, daß dadurch der Verdinglichung des Selbstkonzepts Vorschub geleistet wird, die sowohl von einem behavioristischen Standpunkt (WYLIE 1968) als auch von einem ideologiekritisch-sozialwissenschaftlichen Standpunkt her gesehen unwissenschaftlich erscheint.

Methode der empirischen Erfassung

Die fachbezogenen Selbstkonzepte werden durch die Fragen PVF/PFF-SK.FÄCHER und die auf die Psycho-

logie insgesamt bezogenen durch PVF/PFF-SK.GESAMT bzw. PVF/PFF-SK.ZUKUNFT erfaßt. Die Fragen sind wörtlich im Anhang I wiedergegeben.

Bei der Festlegung der Meßmethode ist vom Ziel der Messung auszugehen. Die Messung soll durch Anregung eines Prozesses geschehen, der eine Bewertung im sozialen Vergleich vornimmt und dabei weitgehend Beurteilungscharakter hat. Der Prozeß soll eine bewußte Einsicht darstellen. Als solcher kann (und sollte) er über eine Operationalisierung erfaßt werden, die direkter Art ist (s. Exkurs I). Anweisung und Formulierungen der entsprechenden Fragen berücksichtigen das.

Als Antwortskala wird eine zehnstufige Skala mit gleichgroßen Prozentrangbereichen verwendet, die jeweils einen Abschnitt durch die Angabe der Percentilgrenzen bezeichnet (s. Anhang I: z.B. PVF/SK.FÄCHER; nach MÜLLER (1975) wäre die Bezeichnung Dezentil-Skala sinnvoll). Es wird vermutet, daß damit der Bewertungsprozeß als möglichst konkreter Vergleichsprozeß provoziert wird. Die Verwendung von Antwortskalen mit Prozentrangbereichen ist in der Forschung zum Selbstkonzept nicht unüblich. Einige Untersuchungen geben Hinweise auf die formale Charakteristik der zu erwartenden Urteile. Es stellt sich heraus, daß bei keiner der Verwendungen eine Verteilung der Antworten im Sinne der durch die Skala nahegelegten Prozentverteilung (in diesem Fall: Gleichverteilung) erreicht wird. Dies gilt sowohl für fünfstufige Prozentskalen mit ungleichen Bereichen (ARSENIAN 1942, Schulische Leistungsfähigkeit) als auch für Quintilskalen (FROEHLICH 1954, mental abilities). Auf Quartilskalen findet ROBERTSON (1960) in einzelnen schulischen Fähigkeiten sowohl Überschätzungen als auch Unterschätzungen. Die berichteten Abweichungen der Antwortverteilungen von dem mit der Skala definierten

Verteilungsmodell und Mittelwertverschiebungen sprechen dafür, daß Bewertungen über Beurteilungen dominieren.

Alternative Methoden sind konkrete Rangbildungen oder einfache Bewertungsskalen (rating-Skalen; OPGENOORTH 1972). Die erste Methode überfordert die Beantworter im vorliegenden Fall; die zweite verzichtet von vornherein auf die Provokation von konkretem sozialem Vergleich als Beurteilung.

In bezug auf die Operationalisierung sind auf Plausibilitätsebene noch einige Einwände aus dem Bereich kritischer Literatur zur Selbstkonzeptforschung zu berücksichtigen. Das betrifft besonders die Mitwirkung der Tendenz zur Abgabe sozial erwünschter Antworten (CROWNE & STEPHENS 1961), die in drei verschiedenen Formen auftreten kann. Als erste ist die Form zu nennen, bei der die verfälschende Tendenz in gleicher Art und Weise die Varianz des Antwortverhaltens der einzelnen Beantworter determiniert. Eine Reihe von Untersuchungen kann eine solche Wirkung bei Selbstkonzeptmessungen nachweisen. Auch die schon erwähnte Verlagerung des Mittelwerts auf Prozentrangskalen ist in diesem Sinn zu deuten. Solange aber die Erwünschtheitstendenz nicht interindividuell unterschiedlich die Antwort beeinflußt (zweite Form) oder sogar selbst interindividuell unterschiedlich ausgeprägt ist (dritte Form), kann keine Verzerrung der Ergebnisse eintreten, zumindest bei den meisten der üblichen Auswertungsmethoden nicht. CROWNE & STEPHENS haben aus vorliegenden Untersuchungen nur Hinweise auf die Wirkung der ersten Form der Erwünschtheitstendenz gefunden.

Darüber hinaus aber ist zu fragen, ob die Wirkung von Erwünschtheitstendenzen überhaupt als Verfälschung des Selbstkonzepts interpretiert werden kann. Unter der Annahme, daß das Selbstkonzept seine

Wirkung immer unter sozialen Bedingungen entfaltet, die ein bestimmtes Ausmaß an Öffentlichkeit mitumfassen, wäre die Erwünschtheitstendenz als "Tendenz zu öffentlicher Selbstdarstellung" integraler Bestandteil des Selbstkonzepts.

COMBS et al. (1963) und WYLIE (1968) listen eine Reihe von weiteren Voraussetzungen auf, die erfüllt sein müssen, damit von einer validen Messung gesprochen werden kann: Nach COMBS et al. muß die "Selbstwahrnehmung" klar sein; für das Selbstkonzept eigener Leistungsfähigkeit ist das eher anzunehmen als für viele andere Selbstkonzepte aus dem weiteren Bereich der Persönlichkeit. Es müssen der mitteilenden Person Ausdruckssymbole verfügbar sein; dafür sind die Skalen vorgesehen. Die Antwortenden müssen zur Kooperation bereit sein; das ist uneingeschränkt der Fall, denn sie können eine Beantwortung ohne persönliche Nachteile ablehnen. Die unterschiedliche Bekanntheit mit den Befragern kann die Mitteilung beeinflussen; Unterschiede in der Bekanntheit aber sind unbedeutend. Erwartungseinstellungen können bestimmte Antworten induzieren; dies gilt insbesondere für die Wiederholungsmessung (PFF), bei der die Beantworter Überlegungen über die Frageabsicht anstellen können. Doch ist die Zeit zwischen den beiden Befragungen mit einigen Wochen so groß, daß eine gezielte Änderung der Antwort des PVF im PFF wenig wahrscheinlich ist.

Validität des Konstrukts im Spiegel empirischer Befunde

Empirische Untersuchungen, die das Konstrukt Selbstkonzept eigener Leistungsfähigkeit verwenden oder ihm zuzuordnen sind, geben in ihrer Gesamtheit ein Bild von der Validität des Konstrukts (wenngleich sie noch keine gezielte "Konstruktvalidierung" darstellen). Unterschiedliche Meßmethoden erschweren zwar

die Zusammenfassung der Befunde, sind aber im Hinblick auf dieses Ziel sogar erwünscht. Bei der Darstellung bleiben Untersuchungen über Bedingungen von Selbstkonzeptänderungen ausgespart (s. Kapitel III.II.).

Eine erste Gruppe von Untersuchungen beschäftigt die Validität und Zuverlässigkeit der Messungen. Die beiden Kriterien lassen sich immer auch auf das begründende Konstrukt beziehen.

Einen geringeren Stellenwert haben Bestimmungen der Zuverlässigkeit, die zudem erstaunlich selten vorliegen. FROEHLICH (1957) fand bei Sekundarschülern eine Übereinstimmung der Selbstbewertung im Abstand einer Woche als Produkt-Moment-Korrelation von .78 für das Selbstkonzept geistiger Fähigkeit. WYLIE (1961) berichtet demgegenüber eine drastische Verringerung der Wiederholungszuverlässigkeit nach elf Wochen Zwischenraum auf einen Wert von r = .20. (Die entsprechende Untersuchung ist leider nicht greifbar.) Sehr hohe Zuverlässigkeiten stellt SHEN (1925) fest, doch bleibt die Methode der Gewinnung unklar.

Die Zahl von Untersuchungen zur Validität dagegen ist groß. Dies rührt u.a. daher, daß ein Vergleich von Selbstbewertungen mit "objektiven Messungen" methodisch einfach durchzuführen ist und vom Grundansatz her naheliegt. Als "objektive" Vergleichsdaten dienen im wesentlichen die Ergebnisse von Intelligenz- und Fähigkeitstests und die Noten in Ausbildungsinstitutionen, seltener auch Fremdbewertungen aus unmittelbarer sozialer Interaktion. (Die Anführungsstriche deuten an, daß eine Diskussion über die Validität der Kriterien selbst notwendig wäre. Setzt man, wie dies hier geschehen soll, eine mäßige Validität der Kriterien voraus, dann bleibt ein weiteres Problem: Ab einer bestimmten Übereinstimmung zwischen Selbstkonzept und Kriterien wird ein Zusammenhang zwischen beiden problematisch, weil

er auf fehlende funktionale Selbständigkeit des Selbstkonzepts hinweist, die auch ein Bestandteil der Validität ist.)

Zusammenfassende Darstellungen von Untersuchungen der Validität von Selbstkonzepten eigener Fähigkeit ergeben eher mäßige Übereinstimmungen mit Kriterien, die bei Korrelationen zwischen .30 und .60 liegen (RUSSELL 1953; WYLIE 1961). In kaum einer Untersuchung wird nicht zumindest ein signifikanter Zusammenhang festgestellt. Für Schüler berichten: AMATORA (1956) eine Korrelation zwischen Intelligenz-Selbstbewertung und dem Urteil der Mitschüler über die Intelligenz von .60, BROOKOVER & THOMAS (1964) eine Korrelation um .40 der Leistungsselbstbewertungen mit dem Notendurchschnitt, SPAIGHTS (1965) überaus hohe Korrelationen bis .85 zwischen den Selbstbewertungen in einzelnen Schulfächern und entsprechenden Testverfahren. BRIM (1966) findet in einer Untersuchung an über 10 000 Schülern in Sekundarschulen eine signifikante Übereinstimmung der Selbstbewertung mit einem Lesetest. Mit dem Alter der Schüler nimmt der Zusammenhang von Selbstkonzept und Testleistung zu (BAILEY & GIBBY 1971), ein möglicher Hinweis auf die Rationalität des Selbstkonzepts.

Der Zusammenhang aber steigt in tertiären Institutionen dann nicht weiter, sondern ist anfänglich gering, um nach Jahren des Besuchs zuzunehmen, was eine rein entwicklungspsychologische Erklärung ausschließt. Die Korrelationen von Noten und Selbstkonzepten für College-Anfänger belaufen sich auf Werte um .20 bis .30, während sie bei älteren Studenten .50 erreichen (BIGGS & TINSLEY 1970; BIGGS & JOHNSON 1972).

Bei allen diesen Ergebnissen ist zu bedenken, daß im amerikanischen Bildungs- und Gesellschaftssystem Testverfahren eine weitaus größere Wertschätzung erfahren und entsprechend mehr gebraucht werden

als etwa in der BRD. Die Zusammenhänge zwischen Noten und Selbstkonzepten haben im tertiären Bereich dort durch die fortlaufende Benotung (begleitende Prüfungen) eine andere Basis.

Unabhängig davon sollten die Ergebnisse zumindest deutlich machen, daß unter bestimmten Bedingungen Selbstkonzepte der Leistungsfähigkeit realistisch sein können, indem sie Übereinstimmung mit "objektiven" Kriterien der Leistung zeigen.

Eine zweite Gruppe von Untersuchungen, die sich mit der Vorhersage von Studienerfolg aus der Selbstbewertung beschäftigen, nimmt eine Zwischenstellung ein. Auf der einen Seite ist sie, wenn ein Zusammenhang gefunden wird, ein weiterer Beweis der Realistik des Selbstkonzepts. Die Ergebnisse der Genauigkeit der Vorhersage des Studienerfolgs schwanken beträchtlich von Untersuchung zu Untersuchung: Bei BIGGS & JOHNSON (1972) signifikante, aber unbedeutende Korrelationen um .20; bei CASHEN (1967) eine Korrelation von .54 für Psychologiestudenten; bei CLARK (1938) eine noch höhere Korrelation von .80 zwischen vorhergesagtem und wirklichem Kurserfolg. Selbst in homogenen Leistungsgruppen sagen Studenten den Kurserfolg mit Korrelationen zwischen .44 und .78 vorher (GAIER 1961). Eine entsprechende Untersuchung an einer deutschen Universität (OPGENOORTH 1972) findet allerdings lediglich eine Übereinstimmung von r = .21 zwischen vorhergesagtem Erfolg in einem Statistikkurs und der wirklichen Note. Abweichend von durchweg signifikanten Zusammenhängen können SIMON & FEATHER (1973) an einem Jahresabschlußexamen keinen entsprechenden Zusammenhang aufweisen.

Auf der anderen Seite aber sind diese Ergebnisse ein erster Schritt zur Bestimmung der funktionalen Validität des Selbstkonzepts, die durch mehr als den Hinweis auf dessen Überlebenswert (GERGEN 1971) zu be-

gründen ist. Das wirkliche Eintreffen vorhergesagten Studienerfolgs kann auch so interpretiert werden, daß die Erwartung den Erfolg selbst bedingt. Die unabhängige Wirkung des Selbstkonzepts aber ist methodisch schwer abzusichern.

Das zeigt sich bei der Untersuchung der Ursachen des Underachievement. Das Selbstkonzept eigener Leistung wird als wesentliches Moment der Verursachung vermutet und in Querschnittsuntersuchungen in dieser Rolle bestätigt (ALEXANDER 1968; BORISLOW 1962), wenngleich Längsschnittuntersuchungen erst volle Klarheit bringen dürften (RAPH et al. 1966).

Eine Versicherung des ursächlichen Einflusses des Selbstkonzepts eigener Leistungsfähigkeit auf den Leistungserfolg will eine Reihe von Untersuchungen durch die Auspartialisierung der Leistungsfähigkeit erreichen. Dabei wird die Annahme gemacht, daß bestimmte Fähigkeitsmessungen alle möglichen vermittelnden Aspekte der Fähigkeit zu kontrollieren in der Lage sind, was insbesondere bei der Auspartialisierung der Fähigkeit über Korrelationen mit Noten fraglich scheint. BIGGS & TINSLEY (1970) stellen eine Beziehung des Selbstkonzepts eigener Leistung zu Noten auch dann fest, wenn die Leistungsfähigkeit in Form eines Leistungstests auspartialisiert ist (partielle Korrelation von .40). Eine weitere Untersuchung kann das Ergebnis nicht replizieren (BIGGS & JOHNSON 1972). Bei BRIM (1954) sinkt die Korrelation auf .20, wenn die Leistung auspartialisiert wird. Der erhaltene Wert aber ist noch signifikant, wie dies auch in der Untersuchung von COOMBS & DAVIES (1966) der Fall ist, wenn die Leistung durch Bildung homogener Gruppen kontrolliert wird.

An der Verwendung dieser Befunde für die Abschätzung der funktionalen Validität des Selbstkonzepts eigener Leistungsfähigkeit ist zu kritisieren, daß da-

mit die in der normativen Betrachtung angestrebte Rationalität verfehlt wird. Eine Funktionalitätsbestimmung dieser Art versucht einzig die Arbeit von BILBY et al. (1972). Die Untersuchung bestimmt den Zusammenhang vom Selbstkonzept eigener Leistungsfähigkeit mit den Ausbildungsplänen bei Schülern 4. bis 6. Schulklassen. Sie stellt dabei signifikante Zusammenhänge von mäßiger Höhe (Korrelationen um .30) fest. Die Enge der Zusammenhänge ist zweifellos am Alter der Versuchspersonen zu relativieren. Die Untersuchung weist auch nach, daß dem Selbstkonzept eine eigenständige Wirkung auf die Pläne zukommt.

Mit dieser Untersuchung ist gleichzeitig der Zusammenhang von Selbstkonzept eigener Leistungsfähigkeit und motivationalen Variablen angesprochen. Auf Konstruktebene wird der folgende Teil D darauf eingehen. Empirische Untersuchungen sind selten und ohne Bezug zueinander. Deshalb muß eine Erwähnung genügen. BAIRD (1970) kann in einer kanonischen Analyse einer großen Stichprobe (über 20 000 College-Studenten) geringfügige Beziehungen von Selbstkonzepten eigener Leistungsfähigkeit zu Lebenszielen nachweisen. DANN & MÜLLER-FOHRBRODT (1972) bestätigen an Oberprimanern eine Beziehung zwischen Berufswünschen und dem Selbstkonzept. Bei derartigen Untersuchungen muß die Frage nach der Kausalität offen bleiben. Der besondere Zusammenhang von Anspruchsniveau (AN) und Selbstkonzept eigener Leistungsfähigkeit ist dort nicht mehr zu untersuchen, wo für beide Variablen gleiche Operationalisierungen verwendet werden (z.B. bei der Vorhersage eigener Leistungen). Wo das AN als konkrete Zielsetzung verstanden wird und das Selbstkonzept als gegenwärtige Bewertung, sind mäßige Zusammenhänge nachweisbar ($r = .46$; GILINSKY 1949); die Beziehungen sind aber eher komplexer Art (WYLIE 1961).

Die bisher ausgewerteten Untersuchungen lassen den Eindruck entstehen, als sei die Validität des Selbstkonzepts, im Sinne der Normwerte realistisch-rationale Selbstbestimmung zu gewährleisten, weitgehend gesichert. Das Bild wird getrübt durch eine weitere große Gruppe von Studien, die Gruppenunterschiede und Über- bzw. Unterschätzungen zum Thema haben. Insgesamt gesehen sind die Ergebnisse dieser Gruppe von Untersuchungen aber deutlich widersprüchlicher als die bisher berichteten. Auch WYLIE (1961) gelingt es nicht, viele der Widersprüche aufzuklären.

Die eindeutigste Situation besteht noch bei den Mittelwertsverschiebungen der Verteilungen der Selbstbewertungen: Für Selbstkonzepte der Leistungsfähigkeit werden durchweg Überschätzungen beobachtet. Das bestätigen die Zusammenfassungen entsprechender Ergebnisse durch RUSSELL (1953) und WYLIE (1961). Im Bereich des Selbstkonzepts eigener Intelligenz werden allerdings auch Unterschätzungen berichtet (ROBERTSON 1960).

Uneinheitlich ist das Bild der vielen Untersuchungen zu Geschlechtsunterschieden: Nach RUSSELL (1953) schätzen sich Mädchen besser ein als Jungen. In den Untersuchungen von BRIM (1966) an Sekundarschülern dagegen ist es umgekehrt. BAILEY & BAILEY (1971) finden bei College-Studenten überhaupt keine Geschlechtsunterschiede, nur auf der Wunschebene zeigen männliche Studenten eine höhere Selbstbewertung. DANN & MÜLLER-FOHRBRODT (1972) können gleichfalls keine Geschlechtsunterschiede bei Oberprimanern feststellen. Nach SCIORTINO (1969) sind in einzelnen Faktoren der Selbstbewertung die Richtungen der Abweichungen unterschiedlich.

Als weiterer Gruppenunterschied wird von BRIM (1966) die soziale Herkunft in die Betrachtung einbezogen. Auch bei Kontrolle der mit der sozialen Herkunft konfundierten Leistung ist das Selbstkonzept von

Schülern mit niedrigerem sozialem Status niedriger.

Der Versuch, die Abhängigkeit der Unter-/Überschätzung der Selbstbewertung von der absoluten Leistungshöhe zu untersuchen, hat bisher nicht zu eindeutigen Ergebnissen geführt. Beobachtungen wie die von CLARK (1938), ROSS (1933) und SUMNER (1932), die gleichsinnig Überschätzungen bei niedrigem Leistungsstand und Unterschätzungen bei hohem Leistungsstand feststellen, können nicht überzeugen, da sie mögliche Regressionseffekte nicht in die Betrachtung einbezogen haben. KEEFER (1971) zufolge sind bessere Studenten zu genaueren Vorhersagen ihrer eigenen Leistungen in der Lage.

Die Betrachtung der Gruppenunterschiede wird noch weiter dadurch kompliziert, daß für bestimmte Gruppen unterschiedliche Validitäten der Selbstkonzepte eigener Leistung möglich sind. Nach BIGGS & JOHNSON (1972) ist die Vorhersage der Leistung aus dem leistungsbezogenen Selbstkonzept bei Studentinnen bedeutend besser möglich als bei Studenten. Dies konnte allerdings von anderen Untersuchungen, wie beispielsweise BROOKOVER & THOMAS (1964) bei Schülern nicht bestätigt werden.

Diese Gruppenunterschiede sind zusammen mit Unter-/Überschätzungen in der durch die normative Begründung des Selbstkonzepts ausgearbeiteten Perspektive Verfälschungen der Realistik. Die vielen (unterschiedlichen) Ergebnisse lassen den Schluß zu, daß solche Verfälschungen zu erwarten sind; ihre Richtung und Ausprägung aber hängen von der konkreten Situation ab, zu der auch die Personengruppe und die Art der Messung zählen. Denn letztlich muß in der Situation auch die Erklärung für die Unterschiedlichkeit der Ergebnisse der berichteten Untersuchungen zu suchen sein.

Zusammenfassung (III.I.C.)

Die verwirrende Vielfalt der Selbst-Begriffe in der amerikanischen Forschung ist Ausgangspunkt des Versuchs einer begrifflichen Klärung. Als grundlegender struktureller Begriff wird der des Selbstkonzepts vorgeschlagen, als dem entsprechende Kennzeichnung prozessualer Art der Begriff der Selbstbewertung.

Die Lösung des Konstruktes Selbstkonzept aus einseitigen Operationalisierungen macht seine Bestimmung als strukturiertes Objekt der Wahrnehmung möglich, das inhaltlich differenziert ist. Eine der Differenzierungen stellt das Konstrukt des Selbstkonzepts eigener Leistungsfähigkeit dar. Es soll den Begriff der Selbstbewertung ablösen, der nunmehr den Prozeß der Konstitution bezeichnet.

Als Inhalt des Selbstkonzepts eigener Leistungsfähigkeit wird die Güte der kognitiven Strukturbildungen in der Psychologie insgesamt (PVF/PFF-SK.GESAMT) wie in den Fächern der einzelnen mündlichen Prüfungen (PVF/PFF-SK.FÄCHER) festgelegt. Als Prozeß wird ein Prozeß der Bewertung unter weitgehender Beteiligung von Beurteilungen angestrebt. Das soll durch die Methode der Abfrage in Form eines Vergleichs mit der Gruppe der Mitprüflinge geschehen. Zusätzlich wird für das auf die Psychologie insgesamt bezogene Selbstkonzept auch noch eine Bewertung durch Einordnung in eine zukünftige Bezugsgruppe verlangt (PVF/PFF-SK.ZUKUNFT). Als Antwortskala wird für den Vergleich eine Skala verwendet, die Prozentränge in zehn gleichgroße Bereiche einteilt. Es kann angenommen werden, daß die Voraussetzungen zur Abgabe wahrheitsgemäßer Urteile erfüllt sind.

Die Auswertung empirischer Untersuchungen zum Selbstkonzept eigener Leistungsfähigkeit ergibt: Die Validität des Konstrukts scheint durch mäßige

Übereinstimmungen mit "objektiven" Messungen von Fähigkeiten einigermaßen gesichert. Auch Untersuchungen über die Vorhersage des Studienerfolgs aus Selbstkonzepten eigener Leistungsfähigkeit bestätigen das. Unabhängig von der Übereinstimmung eignet dem Selbstkonzept funktionale Validität als eigenständige Ursache von Leistungen und Handlungsplanungen.

Eine weitere Gruppe von Untersuchungen läßt aber auch Zweifel an der realistisch-rationalen Bedingung des Selbstkonzepts aufkommen: Es werden häufig bei leistungsbezogenen Selbstkonzepten Überschätzungen beobachtet. Unterschiede zwischen den Geschlechtern, nach sozialer Herkunft und komplexe Unterschiede zwischen verschiedenen Gruppen stellen weitere Einschränkungen der Realistik dar.

III.I.D.: Motivation: Theorie und Methode

Die Darstellung zur Motivation kann und muß wesentlich kürzer ausfallen als die entsprechende Darstellung des Selbstkonzepts eigener Leistungsfähigkeit. Zwei Gründe sprechen dafür: In Kapitel II wurde das Selbstkonzept eigener Fähigkeit eindeutig als die von den Zielen der Untersuchung her gesehen wesentlichere Variable identifiziert. Zweitens ist die Ausgangslage im Verhältnis von Konstruktebene und Operationalisierung bei der Motivation eine andere als beim Selbstkonzept. Dort kann man die Begründung der im Fragebogen vorgenommenen Operationalisierung und die entsprechende Diskussion auf Konstruktebene insofern als synchron bezeichnen, als beide voneinander profitieren können. Eine ganz andere Situation besteht zwischen Konstrukt und Operationalisierung für die Motivation. Auf der Konstruktebene liegen sehr heterogene, weitgehend in Isolation voneinander entwickelte Theorien vor, die entweder auf bestimmte Operationalisie-

rungen festgelegt sind (wie z.B. die Leistungsmotivationsforschung auf das TAT-Verfahren) oder über dem Problem der Operationalisierung schweben (wie z.B. die Theorie der Intrinsischen Motivation).

Aus diesem Grund ist auch die Anordnung der Abschnitte abweichend vom Selbstkonzept: Nach einer kurzen begrifflichen Einleitung werden die Fragen des Fragebogens begründet, die erst anschließend dem empirischen und theoretischen Kontext der Motivationsforschung zugeordnet werden. Dabei gewinnt ein Aspekt besondere Bedeutung: Der Zustand der Motivation der Prüflinge vor der Prüfung, der nicht nur im Hinblick auf die Erfassung der Motivation, sondern auch als mögliche "Einmischung" in das Selbstkonzept eigener Fähigkeiten von Bedeutung sein kann.

Grundcharakter der Motivation

Mit HECKHAUSEN (1965) ist an das Erbe LEWINs anknüpfend Motivation als Person/Umwelt-Bezug zu kennzeichnen. Im Prozeß der Motivation interagieren situative Bedingungen und personale Bedingungen. Die in der Arbeit angesprochenen Rückmeldungen sind mögliche Bestandteile eines solchen Person/Umwelt-Bezuges. In ihnen können bestimmte situative Bedingungen bestimmte personale Bedingungen beeinflussen, die eine Grundlage für die neuerliche Interaktion von Person und Umwelt bilden.

Mit HOLZKAMP-OSTERKAMP (1975) ist Motivation als ein Prozeß emotionaler Wertung zu bestimmen, der zwischen Kognition und Handlung angesiedelt ist. Rückmeldungen stellen Kognitionen über den Zustand von Person/Umwelt-Bezügen dar. Diese Bezüge werden im Prozeß emotionaler Wertung verarbeitet und sind die dynamische Grundlage neuer Person/Umwelt-Bezüge in Form von Handlungen. Die Begriffsverbindung "emotionale Wertung" entspricht jenen allgemeinen Definitionen der

Motivation, die diese als Einheit von energetisierenden und steuernden Prozessen verstehen (COFER & APPLEY 1964; MADSEN 1974).

Der Motivationsbegriff ist für den Gebrauch in dieser Arbeit durch eine gegenständliche Fixierung des Person/Umwelt-Bezuges zu präzisieren. Selbstverständlich ist die Festlegung des Bezuges auf das Umgehen mit wissenschaftlichen Strukturbildungen ("Umgehen" meint nicht weiter differenziert gemäß Kapitel II sowohl die wissenschaftliche Arbeit als auch die Anwendung wissenschaftlicher Erkenntnisse). Als erste Möglichkeit weiterer Festlegung bietet sich der Leistungsaspekt an, der jedoch vom normativen Standpunkt aus als gegenständliche Bestimmung unbefriedigend ist. Beim Selbstkonzept eigener Fähigkeiten wurde eine inhaltliche Kennzeichnung der Strukturbildungen durch die einzelnen Fächer der mündlichen Prüfungen eingeführt. Es liegt nahe, hier gleich zu verfahren. Motivation bedeutet dann die emotionale Wertung von solchen Person/Umwelt-Bezügen, die sich als Studien- und Arbeitsverhalten in den einzelnen Fächern manifestieren. Nun ist sicher ähnlich wie beim Selbstkonzept anzunehmen, daß solche Person/Umwelt-Bezüge z.B. in der Form spezifischer Interessen inhaltlich sehr viel weiter ausdifferenziert sein können und besonders nach dem HD berufsorientiert die Fächereinteilungen hinter sich lassen. Solche individuellen Bezüge jedoch sind nicht miteinander vergleichbar und kaum der beabsichtigten quantitativen Auswertung zugänglich. Die Verwendung der Prüfungsfächer zur gegenständlichen Festlegung der Bezüge macht es zudem wahrscheinlich, mit dieser prüfungsnahen Bestimmung mögliche Effekte der Prüfung nicht zu übersehen.

Eine in vielen Motivationstheorien angelegte grundlegend konflikthafte Auffassung des Motivations-

prozesses, die aktivierende und hemmende Faktoren in ihrer Verflechtung sieht, soll nicht aufgenommen werden. (Die Theorie der Leistungsmotivation zeigt, wie willkürlich ein solches Modell verwendet werden kann, wenn nicht einmal geklärt werden kann, ob es sich bei der "negativen" Komponente um eine Hemmung oder eine andersartige Form der Aktivierung handelt (WEINER 1976).)

Ableitung und Begründung der Fragen
nach Beliebtheit und Handlungsplanung

Der konkrete Inhalt der Fragen zur Motivation ist erst auf dem Hintergrund einer weiteren Eigenart des Motivationsprozesses verständlich. Der Gegenstand der Motivation ist durch eine bestimmte zeitliche Erstreckung gekennzeichnet. Im Bezug zu wissenschaftlichen Strukturbildungen lassen sich aktuelle Vorlieben von solchen Bezügen unterscheiden, die die Grundlage längerfristiger Handlungsstrukturierung bis hin zu ganzen Lebensabschnitten darstellen. Die Unterscheidung deckt sich mit der von LEONTJEW (1973) zwischen Tätigkeitsmotiven und Handlungsmotiven getroffenen.

Der kurzfristigere motivationale Prozeß soll mit der Frage nach der Beliebtheit (PVF/PFF-MOT.BEL.) angesprochen werden. Im PVF geht die Frage allerdings über die reine Zuständlichkeit hinaus, um sich, so weit wie durch eine solche Instruktion möglich, von der Dynamik der konkreten Situation vor der Prüfung (s. unten) zu lösen. Die Frage nach der Beliebtheit der Fächer wird durch den normativen Rahmen nur wenig gestützt. Diesem entspricht die zweite Frage, die sich auf den Einsatz zukünftiger Aktivitäten richtet (PVF/PFF-MOT.W.A.). Sie gilt der (rationalen) Handlungsplanung der Prüflinge, indem sie die geplante Intensität der Weiterarbeit in den einzelnen Fächern erfragt. Dabei bleibt offen, welcher Art im einzelnen

diese Weiterarbeit ist.

Die zweite Frage nach der Handlungsplanung soll dem direkten Nachweis einer Wirkung im Sinne der Normwerte dienen, während die Frage nach der Beliebtheit in bezug auf die Normwerte eher die Funktion einer Kontrollfrage hat: Wirkungen auf die Beliebtheit, die nicht auf die Handlungsplanung sich erstrecken, zeigen einen nur unzureichend "tiefen" Einfluß der Prüfung an. (Wenngleich nicht auszuschließen ist, daß eine von der Beliebtheit geförderte Tätigkeit eine entsprechende Handlungsplanung zur Folge haben kann.)

Die Methode der Fragenformulierung entspricht den Darstellungen von Exkurs I. Beide Fragen werden mit vierstufigen verbal verankerten Antwortskalen versehen und für die einzelnen Fächer der mündlichen Prüfung gestellt. (Ein Bezug zur Psychologie insgesamt wie beim Selbstkonzept ist zumindest für die Frage PVF/PFF-MOT.W.A. wenig sinnvoll und wird deshalb bei beiden Fragen ausgespart.) Die für beide Fragen gleiche Form läßt erwarten, daß mögliche Abweichungen zwischen beiden Aspekten durch Angleichung der Antworten unterdrückt werden können. Die beiden Fragen werden deshalb an unterschiedlichen Stellen der Fragebögen plaziert, wenn das auch eine bewußt vorgenommene Angleichung der Antworten nicht verhindern kann.

Beiträge zur Validität der Motivationsfragen

Dem eklektischen Charakter der beiden Fragen entsprechend stammen Hinweise auf die Validität der Fragen (und die ihnen zugrunde gelegte Begriffsbildung) aus verschiedenen Quellen. Neben Anmerkungen zum Prozeß der Beantwortung stehen begriffliche Vergleiche, einige empirische Arbeiten mit ähnlichen Fragen und abschließend der Versuch, eine Verbindung zu den gängigen Motivationstheorien herzustellen.

Die Situation der Beantwortung der Fragen zur Motivation weicht von der beim Selbstkonzept eigener Leistungsfähigkeit insofern ab, als beim Selbstkonzept eine gewisse Identität des Prozesses des Auskunftgebens und des Prozesses der Selbstbewertung angenommen werden kann, während die Beantwortung der Motivationsfragen keine inhaltliche Beziehung zu der Motivation aufweist, die erfragt wird. Die Frage nach der Intensität der Weiterarbeit (Handlungsplanung) spricht einen zeitlich lang erstreckten Prozeß an, der im Augenblick der Beantwortung nur mehr kognitiv repräsentiert ist. Die Beantwortung ist nicht selbst eine emotionale Wertung. Auch im Falle der Beliebtheit realisiert sich diese als Motivation erst in der aktuellen Beschäftigung mit den einzelnen Gegenständen. Von dieser Warte aus gesehen ist den Antworten zu den Motivationsfragen nicht unmittelbar Validität zuzuerkennen.

Eine empirische Validierung derartiger Messungen kann von der Bestimmung ausgehen, daß Motivationen Vermittlungen zwischen Kognitionen und Handlungen seien. Eine Messung der Motivation erfolgt u.a. in der Absicht, die von der Motivation der Person abhängig gedachten Handlungen besser erklären zu können (Handlungsplanung). Genau das war der Grund für die Einführung des Motivationskonstrukts in die wissenschaftliche Psychologie, als in Experimenten zu viel Varianz unaufgeklärt blieb, wenn allein die aktuelle Situation als Determinante berücksichtigt wurde (COFER & APPLEY 1964). Es sind allerdings nicht allzu enge Beziehungen zwischen Motivation und Verhalten zu erwarten, da die Situation, die motivationale Überdetermination konkreter Handlungen (ATKINSON & FEATHER 1966) und andere Faktoren intervenieren. Dies gilt um so mehr, wenn nicht Handlungen, sondern Handlungsergebnisse als abhängige Variablen verwendet werden.

Mehrere Untersuchungen bestätigen einen Zusammenhang von Motivations- und Interessenmessungen in direkter Abfrage mit Studienverhalten oder Studienerfolg. SCHMIDBAUER (1968) belegt den Zusammenhang von Studienmotivation (berufsmotiviert/fachmotiviert/bildungsmotiviert) mit dem Studienverhalten. Aus zu Beginn des Studiums erfragten Zielen lassen sich die Abschlüsse gut vorhersagen (STAHMANN 1969). Fragebogenmessungen der Leistungsmotivation, die in der Form von Selbsteinschätzungen mit den hier gestellten Fragen formal vergleichbar sind, haben eine Beziehung zum Studienerfolg (im Unterschied zu projektiven Messungen der Leistungsmotivation; BENDIG 1959; LAVIN 1965). Die mögliche Validität von Selbsteinschätzungen der Motivation weist auch die Untersuchung von HOLMES (1971) mit einer Korrelation mit dem Studienerfolg von .54 nach.

Die Hinweise auf empirische Validität direkter Abfrage der Motivation sind nicht befriedigend, weil die verwendeten Methoden im einzelnen doch große Unterschiede aufweisen. Ein Kriterium wie der Studienerfolg ist zudem normativ nicht genügend.

Ein zweiter Schritt der Validierung richtet sich auf die inhaltliche Festlegung auf Beliebtheit und Handlungsplanung. Die rein begriffliche Unterscheidung, die der Begründung der beiden Fragen nach der Motivation dient, treffen ähnlich BEARD et al. (1964) bei der Analyse von Underachievement im akademischen Bereich, wenn sie zwischen intrinsischer Motivation im Sinne eines überdauernden Wertbezuges (die sie durch die Religionsorientierung operationalisieren) und der Beliebtheit als unmittelbarem an die gegenwärtige Umgebung gebundenen Motiv unterscheiden. Beide Konstrukte haben eine voneinander unabhängige Bedeutung für die Erklärung des Underachievement.

Beliebtheit als motivationales Konstrukt (und als Frage) wird von CHAIKIN (1971) in einer experimentellen Untersuchung verwendet. Dort erweist sich diese Variable in einer kurzzeitigen Aufgabe als unabhängig von der Ausdauer.

Der Handlungsplanung als Frage nach der Intensität der Weiterarbeit können die vielgestaltigen Versuche zugeordnet werden, überdauernde Interessen über Fragebögen zu bestimmen, wie dies z.B. für Berufsinteressen geschieht (BAIRD 1970; KUDER 1970). Ein anderer Ansatz, der sich als Konkurrenz zur Leistungsmotivationsforschung versteht, verwendet den Begriff der Intention als eines bewußten Aufgabenbezuges (LOCKE 1966). In Aufgaben von geringer zeitlicher Dauer haben Vornahmen im Sinne konkreter Leistungsziele einen Einfluß auf das Leistungsverhalten. Die Übertragung auf Ziele für längererstrecktes Verhalten im natürlichen Feld, die der Frage nach der Motivation zur Weiterarbeit entsprechen könnte, erfolgt bei LOCKE & BRYAN (1968) durch die Untersuchung der Zielsetzungen in akademischen Prüfungen; solche Zielsetzungen sind eine von der Fähigkeit unabhängige Determination des Erfolgs. (An dieser Stelle taucht aber die Frage auf, ob Zielsetzungen und Motivation begrifflich austauschbar sind; und weiter: wie dazu das Selbstkonzept steht, das auch mit konkreten Zielsetzungen in Teil C in Verbindung gebracht wurde. Der nächste Abschnitt geht auf das Problem ein.)

Rein auf der begrifflichen Ebene kann versucht werden, die eklektischen Begriffe "Beliebtheit" und "Handlungsplanung" in bestehenden Motivationstheorien zu verankern. Als solche kommen die Theorie der Leistungsmotivation und die der Intrinsischen Motivation in Frage.

Die Theorie der Leistungsmotivation zeichnet sich in ihrer orthodoxen Formulierung (z.B. ATKINSON 1964)

durch einige Eigentümlichkeiten aus, die sie in diesem Zusammenhang kaum als brauchbar erscheinen lassen.

1. Die enge Anbindung der Forschung an ein projektives Verfahren (in den USA nur für die positive Tendenz verwendet; im deutschen Bereich für Hoffnung auf Erfolg als auch für Furcht vor Mißerfolg verwendet ; HECKHAUSEN 1963) läßt eine Übertragung der Theorie auf Motivationsmessungen durch Selbsteinschätzungen fragwürdig erscheinen.

2. In Absetzung von McCLELLAND (1966) beschränkt sich die Theorie auf Mikroprozesse des Verhaltens. Darüber hinaus wird Leistungsmotivation, obwohl aus dem Kontext der LEWINschen Psychologie stammend, auf einen in sich geschlossenen Kreis des Verhaltens mit der Motivation am Beginn eingegrenzt, während erst in der letzten Zeit, immer noch als Mikroprozesse, die aus Handlungen folgende Motivation, die hier interessiert, beachtet wird (WEINER 1976).

Vor einer Verwendung der Theorie an zentraler Stelle sollte auch der Hinweis von McCLELLAND (1967) warnen, der die Bedeutung der Leistungsmotivation für wissenschaftliche Tätigkeit gering erachtet. Das schließt nicht aus, die kognitive Wende in der Leistungsmotivationsforschung, die mit Vorgängen wie Selbstbekräftigung und Attribuierung (WEINER et al. 1972) besonders an der Verarbeitung von Rückmeldungen ansetzt, zu nutzen (s. Kapitel III.III.). Als Integrationsansatz jedoch eignet sie sich auch in dieser Form nicht, weil sie den Rahmen, der Grund zur Ablehnung ist, nicht verläßt. Das gilt ebenso für den Versuch von RAYNOR (1970), die Theorie auf Noten in Prüfungen anzuwenden.

Ein begrifflich und vor allem operational weniger festgeschriebener Ansatz ist die Theorie der Intrinsischen Motivation (BERLYNE 1974; PORTELE 1975). In der engeren Fassung setzt sie sich demselben Vor-

wurf der ausschließlichen Beschränkung auf Mikroprozesse aus, wie das bei der Leistungsmotivation der Fall ist. Eine weitere pädagogisch orientierte Verwendung will andeuten, daß Motivation als Person/Umwelt-Bezug in der Form eines unmittelbaren Gegenstandsbezugs möglich und pädagogisch besonders sinnvoll ist. Da jedoch komplexes Verhalten als Ziel der Motivation sicher auch Mittel-Zweck-Relationen beinhalten muß und soziale Zweckbestimmungen ihm wesenseigen sind (ROSENFELD 1966), kann der Ansatz der Intrinsischen Motivation, der als Auflehnung gegen reine Zwangsmotivierungen durchaus brauchbar ist, zur Integration nicht herangezogen werden.

Motivation vor der Prüfung

Bisher wurden Überlegungen zur besonderen Situation der Prüflinge vor der Prüfung und deren Auswirkungen auf die Validität der Antworten ausgespart. Es liegt nahe, diese Einflüsse der Situation auf die Antwort zur Frage nach dem Selbstkonzept eigener Leistungsfähigkeit über die Motivation der Prüflinge vermittelt zu sehen. Damit wird zugleich ein Beitrag zur weiteren Validierung der Antworten zu den Motivationsfragen geleistet.

Die Einflüsse der besonderen Situation vor der Prüfung auf die Antworten können in zwei Anteile aufgespalten werden: Einflüsse der Prüfungsvorbereitung bis zur Beantwortung und Einflüsse der erwarteten Prüfung. Die Einflüsse der Prüfungsvorbereitung stehen nicht zur Diskussion (s. Teil A), sondern nur die erwartungsbedingten (zukunftsperspektivischen) Einflüsse. Ziel der Untersuchung muß es sein, den Zustand der Prüflinge im festgelegten zeitlichen Abstand vor der Prüfung (ungefähr eine Woche) möglichst frei von diesen Erwartungen zu erfassen.

Ein erster Schritt ist die Analyse des motiva-

tionalen Zustands der Prüflinge vor der Prüfung. FISCH (1970) kann vor Psychologieprüfungen bedeutsame Änderungen der Leistungsmotivation feststellen. Dabei ist gerade zu der Zeit, zu der dieser Fragebogen beantwortet werden soll, ein krisenhafter Zustand erreicht. Eine Untersuchung von PIEHL (1972), die gleichfalls bei Psychologiestudenten durchgeführt wurde, findet bedeutsame Aktivierungsänderungen allerdings besonders kurz vor der Prüfung (1 Tag). LUNDBERG & EKMAN (1970) berichten einen exponentiellen Anstieg der Erregung mit verringertem zeitlichen Abstand zur Prüfung. Die Beschleunigung wird dabei noch zusätzlich durch die Abhängigkeit dieses Anstiegs von der subjektiven Zeit beeinflußt, die bei zunehmender Annäherung der Prüfung immer mehr unterschätzt wird. Da die Ergebnisse von FISCH (1970) aus Rückerinnerungen gewonnen werden, ist den beiden anderen Untersuchungen mehr Gewicht einzuräumen und sind deshalb weniger gravierende Änderungen im einwöchigen Abstand anzunehmen.

Die Auswirkungen sind nicht nur quantitativer, sondern auch qualitativer Art. Konkurrenzbezogene Notensysteme vermögen zwar, zu Leistungen zu aktivieren, das Interesse und die Zufriedenheit jedoch können negativ beeinflußt werden (COLEMAN 1959; DEUTSCH 1951). Untersuchungen über Notensysteme, die nur ein Urteil nach Bestanden/Nicht-Bestanden fällen, zeigen entsprechend eher leistungsverringernde Wirkung (CLARK 1969; KARLINS et al. 1969), während die Zufriedenheit der Studenten in solchen Systemen steigt; die Befunde sind aber nicht immer übereinstimmend (MANELLO 1964). (In Kap. III.II. wird das als handlungsüberdauernde motivationale Wirkung der Prüfung noch einmal aufgenommen.)

Die Antworten auf die Motivationsfragen könnten mehr von den im letzten Abschnitt beschriebenen Auswir-

kungen betroffen werden als von den rein quantitativen Aktivierungen, da sie die Leistungsthematik, die den Aktivierungen zugeordnet ist, nicht unmittelbar ansprechen. Für die Frage nach der Beliebtheit (PVF-MOT.BEL.) versucht die Instruktion, durch den Bezug zum Studium die Wirkung der Situation zu kontrollieren, wenngleich auch bei direkten Fragen eine die Kontrolle verwirklichende Rationalität der Antwort nicht gesichert ist.

Die situative Abhängigkeit der Antworten wird davon abhängen, inwieweit die Befragung eine anschauliche Antizipation der erwarteten Prüfung erzeugt. FISCH (1970) weist die aktivierende Wirksamkeit von konkreten Hinweisreizen auf Prüfungen nach. Der Fragebogen als direkte Befragung ist darauf angewiesen, die kommende Prüfung anzusprechen; das erhöht die Wahrscheinlichkeit von nicht erwünschten Auswirkungen zukunftsperspektivischer Art, ist zugleich aber auch Möglichkeit ihrer bedingten Kontrolle durch Aufforderungen zum kognitiv kontrollierten Bericht.

Mögliche Wirkungen auf die Antworten zum Selbstkonzept eigener Fähigkeiten werfen erneut die Frage nach der Beziehung von Selbstkonzept und Motivation auf. MEYER (1973) interpretiert die Leistungsmotivation als Selbstkonzept eigener Begabung. Aber diese rein begriffliche Analyse, die den Bereich der Selbstkonzeptforschung nicht integriert, löst in der Tradition der kognitiven Wende die Eigenart motivationaler Prozesse so weit auf, daß sie das Selbstkonzept eigener Begabung über ein "Informationsbestreben" mit dynamischer Qualität ausstatten muß. Es ist weniger spekulativ, eine Verbindung von Selbstkonzept und Motivation unter leistungsthematischem Aspekt über das Anspruchsniveau herzustellen. Eine einfache Gleichsetzung von Selbstbewertung und Anspruchsniveau (ISRAEL 1976) ist zwar voreilig, doch zeigten schon die Befunde in Teil C, daß Vorhersagen eigener Leistungen in den Kon-

text des Selbstkonzepts eigener Fähigkeiten passen. Das Anspruchsniveau läßt sich als Erwartung eigener Leistung (s. Kapitel III.III.) als solche Vorhersage operationalisieren. Wird auch an späterer Stelle der Versuch unternommen, die konkrete Vorhersage als Antizipation von der zuständlichen Selbstbewertung abzutrennen, dann schließt doch die konkrete Situation mit der Aktivierung leistungsthematischer Bezüge nicht aus, daß die Antworter die Frage nach dem Selbstkonzept zukunftsperspektivisch motivational akzentuiert beantworten. Durch personal unterschiedlich verfestigte Motivationsstrukturen könnte das die Validität des Selbstkonzepts eigener Leistungsfähigkeit im PVF verringern.

Die Analyse sagt wenig über das empirische Verhältnis von Fragen zur Motivation und zum Selbstkonzept im Fragebogen aus, da die Motivationsfragen nicht dem leistungsthematischen Aspekt der Motivation zuzuordnen sind (s. oben).

Zusammenfassung (III.I.D.)

Der Begriff der Motivation, der den Fragen nach der Handlungsplanung (PVF/PFF-MOT.W.A.) und nach der Beliebtheit (PVF/PFF-MOT.BEL.) zugrunde liegt, zeichnet sich seinem Grundcharakter nach als emotionale Wertung und als Person/Umwelt-Bezug aus. Inhalt und Gegenstand des Bezuges sind Studien- und Arbeitsverhalten in den einzelnen Fächern der Prüfungsordnung.

Die Frage nach der Handlungsplanung ist unmittelbar zielbezogen (Kapitel II), während die Frage nach der Beliebtheit eher eine Kontrollfrage darstellt.

Zur Einschätzung der Validität der Fragen tragen verschiedene Quellen bei. Eine Analyse der Situation macht deutlich, daß die Beantwortung der Fragen selbst

noch keinen motivationalen Prozeß im Sinne der Fragen darstellt. Empirische Untersuchungen zeigen, daß sich aus direkt erfragter Motivation Studienverhalten und Studienerfolg vorhersagen lassen. Die angesprochenen Prozesse selbst finden auf begrifflicher und empirischer Ebene aus anderen Untersuchungen einige Unterstützung. Der Versuch, die Begriffe in bestehenden Motivationstheorien, der Theorie der Leistungsmotivation und der Theorie der Intrinsischen Motivation, zu verankern, ist nicht erfolgversprechend.

Abschließend wird der unerwünschte zukunftsperspektivische Einfluß der Prüfung, der über die Motivation der Prüflinge vermittelt auf die Beantwortung der thematischen Fragen einen Einfluß nehmen kann, erörtert. Eine Beeinflussung des Selbstkonzepts durch die leistungsthematische Aktivierung vor der Prüfung kann nicht ausgeschlossen werden, ist aber eine Woche vor der Prüfung eher gering. Auswirkungen auf die Motivation (als Verringerung der angesprochenen Motivation) sind möglich, aber von den Antwortern kognitiv kontrollierbar.

III.I.E.: Empirische Eigenschaften von Selbstkonzept und Motivation

Die Auswertung der Daten im Rahmen der Voraussetzungsanalyse kann nur einige Aspekte, die in den theoretischen Teilen C und D erwähnt wurden, mit ihren Mitteln angehen, denn zur Bearbeitung vieler Probleme wären eigenständige Untersuchungen notwendig. Ohnehin ist das empirische Vorgehen der Voraussetzungsanalyse dadurch belastet, daß dieselben Daten die Grundlage der inhaltlichen Auswertung darstellen. Wenn empirische Validierungen der Antworten angestrebt werden, sind sie von den Ergebnissen selbst nicht unabhängig. Dies gilt auch für die Methode der Darstellung: Die Ergebnisse der Voraussetzungsanalyse enthalten immer

schon Hinweise auf die Ergebnisse der folgenden Auswertungsschritte. Diese Einwände sowie die Stellung des PVF als "neutraler" Bezugspunkt und die Zweifel an dieser Neutralität im vorangehenden Teil sind Anlaß, die Auswertung so weit wie möglich auf die Daten des PVF zu beschränken.

Übersicht: Die Auswertung geht in drei Schritten vor.

Im Teil 1 werden die Eigenschaften der Antwortverteilungen zum Selbstkonzept eigener Leistungsfähigkeit auf Verteilungsform und Mittelwertverschiebungen hin untersucht. Die Antwortverteilungen für Beliebtheit und Handlungsplanung werden nicht einbezogen, da eine entsprechende Auswertung schon in Exkurs I stattfand.

Die Abhängigkeit der thematischen Fragen von bestimmten Bedingungen personaler und situativer Art ist Gegenstand von Teil 2 und soll Hinweise auf die Validität der Antworten geben.

Dem gilt auch die Auswertung der Zusammenhänge (Korrelationen) zwischen den thematischen Fragen im 3. Teil.

Die Auswertung der auf die einzelnen Fächer bezogenen Antworten erfolgt auf den in Exkurs II beschriebenen Ebenen, teilweise aber auch davon abweichend in den einzelnen Fächern. Zusätzlich werden bei den auf die Psychologie insgesamt bezogenen Fragen zum Selbstkonzept "normale" Auswertungen über die Prüflinge hinweg vorgenommen.

III.I.E.1.: Analyse der Antwortverteilungen zum Selbstkonzept

Die Analyse der Verteilungen der Selbstkonzepte eigener Fähigkeit kann Aufschlüsse über den Prozeß der Beantwortung geben. Die Auswertung bezieht die Ergeb-

nisse beim PFF mit ein, da für diese Fragen nicht wie für die übrigen Fragen mit Antwortskalen eine Prüfung auf Normalverteilung (als Grundlage der statistischen Auswertung) bereits in Exkurs I dargestellt wurde.

Die Instruktion zur Frage nach dem Selbstkonzept legt einen Prozeß konkreten Vergleichs nahe. Wäre der so möglich und würde die Beurteilung eigener Fähigkeit realistisch erfolgen, dann müßte die Verteilung der Antworten die Form einer Gleichverteilung aufweisen. Die in Teil C berichteten Ergebnisse bisheriger Untersuchungen lassen daran zweifeln.

Tabelle: III.I. - 12

Verteilungsparameter und Prüfung der Verteilungsform für PVF/PFF-SK.GESAMT/ZUKUNFT

Fragen	Wahrscheinlichkeit der Abweichung von		Nur PVF:		Wahrscheinlichk.zuf. Abweich. von
	Gleichverteilung	Normalverteilung	arithm. Mittel	Stand. Abw.	AM = 4.50
Vordiplom					
PVF-SK.GESAMT	<.05	o.A.	5.48	1.42	<.01
PFF-SK.GESAMT	<.05	o.A.			
PVF-SK.ZUKUNFT	<.01	o.A.	6.10	1.56	<.01
PFF-SK.ZUKUNFT	<.01	o.A.			
Hauptdiplom					
PVF-SK.GESAMT	<.05	o.A.	6.33	1.78	<.01
PFF-SK.GESAMT	<.01	o.A.			
PVF-SK.ZUKUNFT	<.05	o.A.	5.74	1.83	<.01
PFF-SK.ZUKUNFT	<.01	o.A.			

Methodisches Vorgehen: Die Überprüfung auf Güte der Anpassung erfolgt mit dem Kolmogoroff-Smirnoff-Anpassungstest (LIENERT 1973). Wegen der deskriptiven Bedeutung sind Signifikanzangaben bis zu p = .20 angegeben; liegen die Wahrscheinlichkeiten darüber, ist ein "o.A." vermerkt. Die Überprüfung der Mittelwerts-

abweichungen vom "theoretischen" Mittelwert wird mit dem t-Test vorgenommen. Ein höherer Mittelwert entspricht einem höheren Selbstkonzept.

Die Verteilungsparameter und Überprüfungen der Mittelwertverschiebungen werden, um die Vorwegnahme von Ergebnissen von Kapitel III.II. zu umgehen, nur für den PVF angegeben.

Die Auswertung der Antworten zu den einzelnen Fächern wird nicht in den "Fach-Prüfer-Kombinationen", sondern in den einzelnen Fächern vorgenommen. Dem liegt die Vermutung zugrunde, daß die teilweise geringen Umfänge einzelner "Fach-Prüfer-Kombinationen" die Erkennbarkeit von charakteristischen Eigenschaften der Antwortverteilungen erschweren, die Zusammenfassungen in einem Prüfungsfach aber solche Eigenschaften kaum verdecken dürften.

Tabelle: III.I. - 13

Verteilungsparameter und Prüfung der Verteilungsform für PVF/PFF-SK.FÄCHER

	Wahrscheinlichkeit der Abweichung von				Nur PVF:		
	Gleichverteilung		Normalverteilung				
	PVF	PFF	PVF	PFF	arithm. Mittel	Stand. Abw.	Wahrscheinlichk.zuf. Abweich. von AM = 4.50
Vordiplom							
Ausdr.	o.A.	$\leq$.10	o.A.	o.A.	4.90	2.37	
Method.	o.A.	$\leq$.05	o.A.	o.A.			
Allg.Psychol.	$\leq$.05	$\leq$.01	o.A.	o.A.	5.47	1.72	$\leq$.01
Entw.Psychol.	$\leq$.05	$\leq$.05	o.A.	o.A.	5.56	1.92	$\leq$.01
Pers.Psychol.	$\leq$.20	$\leq$.01	o.A.	o.A.	5.45	1.93	$\leq$.01
Physiol.	$\leq$.20	$\leq$.05	o.A.	o.A.	4.79	2.19	
Wahlfach I	o.A.	$\leq$.01	o.A.	o.A.	5.31	2.08	$\leq$.05
Wahlfach II	$\leq$.10	$\leq$.05	o.A.	o.A.	5.82	1.61	$\leq$.01

Tabelle: III.I. - 13 (Fortsetzung)

Verteilungsparameter und Prüfung der Verteilungsform für PVF/PFF-SK.FÄCHER

	Wahrscheinlichkeit der Abweichung von Gleichverteilung PVF	Gleichverteilung PFF	Normalverteilung PVF	Normalverteilung PFF	arithm. Mittel	Stand. Abw.	Nur PVF: Wahrscheinlichk.zuf. Abweich. von AM = 4.50
Hauptdiplom							
Diagnostik	<.20	<.05	o.A.	o.A.	5.52	2.11	<.05
Sozialpsychol.	o.A.	<.01	o.A.	o.A.	5.17	2.12	
Päd.Psychol.	<.05	<.01	o.A.	o.A.	5.83	1.91	<.01
Klin.Psychol.	<.01	<.01	o.A.	o.A.	6.45	1.82	<.01
Angew.Psychol.	o.A.	<.05	o.A.	o.A.	4.72	2.20	

Methodisches Vorgehen: s. Tabelle: III.I. - 12. Ein höherer Mittelwert entspricht einem höheren Selbstkonzept. Für das erste Fach des VD werden die Prüflinge der NPO und APO getrennt ausgewertet.

Die Ergebnisse lassen im Hinblick auf die statistische Auswertung die Anwendung von Verfahren, die Normalverteilungen voraussetzen, unproblematisch erscheinen.

Die inhaltliche Auswertung der Tabellen: III.I. - 12/13 bezieht neben der Verteilungsform auch die Mittelwertverschiebungen der Verteilungen ein.

Die Verteilungsform aller Antwortverteilungen zum Selbstkonzept läßt selbst dann keine Abweichung von der Normalverteilung erkennen, wenn eine statistische Fehlergrenze von p = .20 gesetzt wird. Demgegenüber weichen die meisten Verteilungen statistisch bedeutsam von dem Modell der Gleichverteilung ab. Daß einige Antwortverteilungen weder Abweichungen von der Normalverteilung noch von der Gleichverteilung zeigen, ist bei Stichproben derart kleinen zahlenmäßigen Umfangs kein Widerspruch.

Die deutliche Abweichung der Antwortverteilungen von der Gleichverteilung und ihre weitgehende Übereinstimmung mit der Normalverteilung läßt erkennen, daß sich der Antwortprozeß vom Ergebnis her gesehen nicht von Bewertungsprozessen unterscheidet, die auf rating-Skalen vorgenommen werden. Die Beantwortung kann nicht die Rationalität beanspruchen, die ein konkreter Vergleich mit einer Bezugsgruppe hätte. Die Antworten erfolgen mehr im Sinne von globalen Bewertungen, wie das in Teil C bereits angedeutet wurde. Der Grund dafür ist aber nicht unbedingt in "irrationalen" Momenten der Antwortsituation wie z.B. "Einmischungen" von einer leistungsthematischen Perspektive zu suchen, sondern die Möglichkeiten konkreten Vergleichs sind von vornherein begrenzt. Den Prüflingen fehlen Informationen über die Leistungen ihrer Mitprüflinge, die einen konkreten Vergleich erlauben würden.

Es ist zu fragen, ob das Selbstkonzept, das in solchen globalen Bewertungsprozessen konstituiert wird, im Sinn der Untersuchungsabsicht genügend Rationalität aufweist. Die Rationalität des Selbstkonzepts ist in seinem Beitrag zu Entscheidungen zu sehen, die selbst kaum konkrete Vergleichsgruppen und noch weniger diese konkrete Vergleichsgruppe der Mitprüflinge voraussetzen (wie z.B. Entscheidungen für die Berufswahl, die weitere Ausbildung usw.). Mit dieser Auffassung entsteht allerdings ein neues Hindernis zur Wirkung der Noten auf das Selbstkonzept, da die Informationsentnahme bei den Noten konkret vergleichsbezogen sich empfiehlt (Teil B).

Aus den Tabellen: III.I. - 12/13 wird ein weiteres Merkmal der Antwortverteilungen deutlich, das gleichfalls in anderen Untersuchungen anzutreffen ist: die Verschiebung der Mittelwerte der Verteilungen zum positiven Pol der Antwortskala hin. Der Mittelwert der Antworten müßte der Verschlüsselung zufolge bei 4.50

liegen. Mit einigen Ausnahmen einzelner Fächer sind die Mittelwerte statistisch bedeutsam "besser". Im HD könnte das auf die Folgen (oder Ursachen, s. unten) der Selbstselektion zurückgeführt werden; für das VD aber ist diese Erklärung nicht möglich und deshalb auch für das HD weniger wahrscheinlich zutreffend.

Die Verlagerung der Mittelwerte ergibt sich nicht schon unmittelbar aus der Normalverteilung der Antworten. Sie muß im Zusammenhang mit den in Teil C dargestellten Untersuchungen gesehen werden, die in der Mehrzahl bei allgemein positiv gewerteten Eigenschaften, wie sie Intelligenz und Leistungsfähigkeit darstellen, Überschätzungen berichten. Sich in Mittelwertverschiebungen äußernde Überschätzungen werden gewöhnlich auf Tendenzen zur Abgabe einer sozial erwünschten Antwort zurückgeführt. Da bei dieser Messung des Selbstkonzepts aber eine gewisse Identität von Meßprozeß und realem Prozeß besteht, ist an Stelle der methodischen Interpretation eine inhaltliche durchaus angebracht: Die Ergebnisse lassen den Schluß zu, daß die Realistik des Selbstkonzepts eigener Fähigkeit durch eine Tendenz zur Überschätzung eingeschränkt wird.

Zusammenfassung (III.I.E.1.)

Sowohl die auf die Psychologie insgesamt bezogenen Selbstkonzepte als auch die fachbezogenen Selbstkonzepte eigener Leistungsfähigkeit weisen als Antwortverteilungen eher Normalverteilungen als Gleichverteilungen auf. Sie können deshalb nicht die Rationalität eines konkreten Vergleichs für sich beanspruchen, was aber für ihre praktische Bedeutung auch nicht notwendig ist. Von wenigen Ausnahmen abgesehen sind die Mittelwerte der Antwortverteilungen statistisch bedeutsam zum positiven Pol der Skala hin verschoben. Dies deutet eine Einschränkung der Realistik des Selbstkonzepts durch Überschätzung an.

III.I.E.2.: Persönlichkeits- und Situationsbedingungen

Die Analyse von Bedingungen des Antwortverhaltens in den thematischen Fragen dient vorrangig der Validierung. Darüber hinaus können die Ergebnisse auch einen Hinweis darauf geben, ob in der Wirkungsanalyse bedingungsspezifische Effekte zu erwarten sind.

Unterschiede zwischen den Prüfungsarten (VD und HD)

Die Auswertung der Unterschiede von VD und HD wird durch die an der Verteilung der Noten festgestellte Selbstselektion von Prüflingen mit schlechten Noten (Exkurs I) erschwert: Es liegt nahe, Selbstkonzept und Motivation als Ursachen der Selbstselektion, die im wesentlichen schon beim Ausfüllen des PVF (also vor Kenntnis der Noten) stattfindet, beteiligt zu sehen.

Tabelle: III.I. - 14

Unterschiede der Verteilungsparameter zwischen VD und HD (nur PVF)

	arithmetische Mittelwerte		Wahrscheinlichk.zuf. unterschiedl.	
	VD	HD	Varianz	Mittelwerte
Gesamtkonzepte				
PVF-SK.GESAMT	5.48	6.33	n.s.	<.05
PVF-SK.ZUKUNFT	6.10	5.74	n.s.	n.s.
Werte Über-alle-Fächer				
PVF-SK.FÄCHER	5.32	5.54	n.s. (?)	n.s. (?)
PVF-MOT.BEL.	2.40	2.20	n.s. (?)	<.05 (?)
PVF-MOT.W.A.	2.90	2.37	n.s. (?)	<.01 (?)

Methodisches Vorgehen: Die Varianzen der beiden zu vergleichenden Antwortverteilungen werden mit dem F-Test auf Gleichheit überprüft. Da in keinem Fall si-

gnifikante Unterschiede beider Varianzen festgestellt werden können, werden die Mittelwerte mit einem t-Test für homogene Varianzen auf Unterschiedlichkeit geprüft. Ein höherer Mittelwert entspricht einem höheren Selbstkonzept und minderer Motivation.

Die statistische Prüfung der Werte "Über-alle-Fächer" ist, wie schon mehrfach betont, mit einem (?) versehen, weil die Unabhängigkeit der einzelnen Antworten einer Person nicht gesichert ist.

Das signifikant höhere gegenwartsbezogene Gesamtkonzept (im folgenden als Kürzel für "Selbstkonzept in der gesamten Psychologie" verwendet) bei den Prüflingen des HD könnte personenspezifisch auf die Selbstselektion zurückgeführt werden. Eine derartige Erklärung müßte einen ähnlichen Unterschied für das zukunftsbezogene Selbstkonzept behaupten. Dort aber weisen die Prüflinge des HD im Mittel (nicht signifikant) niedrigere Selbstbewertungen auf. Diese Diskrepanz erfordert eine situative Erklärung. Die Unterschiede zwischen beiden Gesamtkonzepten könnten mit dem Begriff der Kompetenz erklärt werden. Während im Studium vor dem VD bei großer Stoffülle und größerer Anzahl der Einzelfächer nur ein vergleichsweise grober Überblick über die einzelnen Prüfungsfächer erreicht werden kann, sind die Bedingungen nach einem weiteren Studienabschnitt und bei weniger Fächern im Hauptdiplomstudium besser beschaffen. Hinzu tritt für das Studium vor dem HD ein Mehr an Praxisbezug, der zusätzliche Kompetenzerlebnisse ermöglicht. Die Deutung wird durch die Ergebnisse von ALVES & LODDERS (1974) abgesichert, die im HD ein deutlich größeres Ausmaß an selbstbeurteilter Beherrschung der Fächer feststellen konnten.

Das Urteil eigener Kompetenz für Zukunftsleistungen ist abhängig von der Klarheit der Anforderungen. Im VD sind die Anforderungen des HD klar erkennbar, während die beruflichen Anforderungen, die als Urteils-

grundlage im PVF des HD verlangt werden, kaum entsprechend klar sein dürften. Das veranlaßt die Prüflinge des HD zur vorsichtigeren Selbstbewertung.

Die Interpretation impliziert die Annahme, daß die Antwort zum Selbstkonzept mehr an absoluten Bezugssystemen bewertend (Kompetenz) als vergleichsorientiert erfolgt, und paßt sich damit in die Tendenz der Interpretation der Verteilungseigenschaften ein. Die Unterschiede zwischen den beiden Gesamtkonzepten sind ein erster Hinweis auf die differentielle Validität beider Fragen.

Die Ergebnisse der Auswertung "Über-alle-Fächer" schränkt die Deutung der Unterschiede der Gesamtkonzepte mit Hilfe des Kompetenzbegriffs ein, denn die Unterschiede in der erlebten Kompetenz müßten sich ähnlich auch in den einzelnen Fächern abzeichnen. Die Mittelwerte des HD sind aber nur wenig höher als die des VD.

Demgegenüber beantworten die Prüflinge des HD beide Motivationsfragen mit "positiveren" Antworten; bei der Handlungsplanung ist der Unterschied sehr viel größer als bei der Beliebtheit. Sieht man von möglichen Einflüssen der Selbstselektion ab, die kaum die Unterschiede zwischen den beiden Fragen erklärt, ist die "höhere" Motivation auf dem Hintergrund der Studien- und Arbeitsperspektive der beiden Studentengruppen verständlich. Das Studium nach dem VD verlangt die Umstellung auf ein anderes Fächerangebot; die Antworten auf die Frage nach Intensität der Weiterarbeit (Handlungsplanung) in den Fächern des VD liegen dementsprechend in der Gesamttendenz in der Nähe der verbalen Skalenverankerung "wenig intensiv". Ein großer Teil der Studenten, die nach dem HD praktisch tätig werden, wie auch ein Teil derjenigen, die das Studium fortsetzen, werden dagegen in den Bereichen weiterarbeiten, die von den Fächern des HD abgedeckt werden. Die deutliche Verlagerung des Mittelwerts

aller Antworten zur verbalen Skalenverankerung "intensiv" entspricht dieser Situation. Sie ist zugleich eine Konstruktvalidierung der Frage.

Für die unterschiedliche Beliebtheit im Studium (PVF) liegt es nahe, unter Berücksichtigung des geringeren Ausmaßes des Unterschiedes zwischen VD und HD, die Beliebtheit als Folge der Handlungsplanung zu betrachten. Eine Alternativerklärung wäre: Die intensivere Auseinandersetzung mit den Fächern des HD erlaubt mehr kognitive Konflikte (BERLYNE 1974), Kompetenzerlebnisse usw.; das trägt zu einer Erhöhung der Beliebtheit bei. Eine Klärung ist im Rahmen dieser Auswertung nicht möglich. Damit ist zugleich eine differentielle Validität beider Fragen nicht eindeutig erwiesen.

Unterschiede zwischen den "Fach-Prüfer-Kombinationen"

Die "Fach-Prüfer-Kombinationen" stellen eine wesentliche Ebene der Auswertung dar. Sie lassen sowohl Unterschiede zwischen den Fächern als auch zwischen den Prüfern deutlich werden; vermittelt können Unterschiede auch durch die Prüferwahl bedingt sein.

Grob gesehen geben bedeutsame Mittelwertdifferenzen zwischen ihnen beim Selbstkonzept Hinweise auf nicht rationale Quellen der vergleichenden Selbstbewertung (von Einflüssen der Prüferwahl abgesehen), während solche Differenzen für die Motivationsfragen, besonders die Handlungsplanung, auf Grund der dominanten Berufsperspektiven für das HD und der Ausbildungsstruktur des Hauptdiplomstudiums zusammen mit der Berufsperspektive für das VD zu erwarten sind. Die Auswertungen erfolgen wiederum nur an den Antworten des PVF.

Tabelle: III.I. - 15

Mittelwertunterschiede in den "Fach-Prüfer-Kombinationen" (nur PVF)

	Wahrscheinlichk. heterogener Varianzen		zuf. unterschiedl. Mittelwerte	Eta-quadrat
	Bartlett	Cochran	F-Test	
Vordiplom				
PVF-SK.FÄCHER	n.s.	n.s.	= .67 (?)	-
PVF-MOT.BEL.	n.s.	n.s.	<.01 (?)	.17
PVF-MOT.W.A.	<.01	<.01	<.01 (?)	.40
Hauptdiplom				
PVF-SK.FÄCHER	n.s.	n.s.	<.01 (?)	.16
PVF-MOT.BEL.	n.s.	n.s.	<.05 (?)	.13
PVF-MOT.W.A.	n.s.	n.s.	<.01 (?)	.18

Methodisches Vorgehen: Die Mittelwertunterschiede der Antwortverteilungen in den einzelnen "Fach-Prüfer-Kombinationen" werden mittels einer einfachen Varianzanalyse überprüft. Dabei wird angenommen, daß die einzelnen Antworten voneinander unabhängig sind. Auch hier ist die Einschränkung zu machen, daß zwischen den einzelnen "Fach-Prüfer-Kombinationen" im Ausmaß der Überschneidung durch identische Prüflinge Abhängigkeiten vorliegen können. Die Richtung möglicher Fehleinschätzung von Signifikanzen aber ist nicht anzugeben. Voraussetzung der Varianzanalyse ist die Homogenität der Varianzen in den einzelnen Gruppen (Kombinationen). Mit einer Ausnahme ist diese Annahme der Varianzhomogenität erfüllt (s. Spalte 1 und 2 der Tabelle). Da in diesem Fall ein besonders hoher F-Wert (>16 bei df = 10/261) vorliegt, kann die Abweichung übergangen werden. Die Eta-Quadrat-Werte der letzten Spalte werden im laufenden Text erläutert.

Im Selbstkonzept eigener Leistungsfähigkeit sind für die "Fach-Prüfer-Kombinationen" des VD keine be-

deutsamen Mittelwertunterschiede festzustellen. Die Antworten auf die beiden Motivationsfragen dagegen zeigen solche Unterschiede. Für die Prüflinge des HD sind gleichfalls bedeutsame Mittelwertunterschiede in den Motivationsfragen, zusätzlich aber auch ein Unterschied im Selbstkonzept nachzuweisen.

Im Sinne der anfänglichen Erwartungen sind die Ergebnisse überwiegend Validierungen der Fragen. Dies kann und muß noch konkretisiert werden. Die Frage nach dem Selbstkonzept verlangt einen Vergleich der eigenen Fähigkeit mit der der anderen Prüflinge. Unterschiede zwischen den Fächern dürfte es nicht geben, da die Vergleiche innerhalb der Fächer stattfinden sollen. Unterschiede zwischen Prüfern sind aber nicht ganz ausgeschlossen: Sie könnten auf die unterschiedliche Vorbereitung für einzelne Prüfer (Erwerb von Fähigkeiten) und die aktive Auswahl der Prüfer durch die Prüflinge zurückgehen. Solche Unterschiede müßten aber insgesamt gering bleiben. Die bisherigen Ergebnisse deuten noch eine weitere Möglichkeit an, wenn das Selbstkonzept vom Erleben eigener Kompetenz beeinflußt wird (was aber nur für die Gesamtkonzepte als Interpretation nahegelegt wurde): Fächer können durch unterschiedliche Ausbildungsbedingungen ein unterschiedliches Ausmaß an Kompetenzerleben ermöglichen und so zu fachgebundenen Differenzen beitragen.

Die Motivationsfragen verlangen ein absolutes Urteil. Durch die unterschiedliche Gewichtung der Einzelfächer in der Ausbildung und die konkreten Schwerpunkte in der Studien- und Berufsperspektive bedingt, sind Unterschiede zwischen Fächern sinnvoll. Unterschiede zwischen den Prüfern sind nicht ganz auszuschließen, weil auch die Motivation an der Auswahl der Prüfer mitwirken kann; das Ausmaß sollte aber vergleichsweise gering sein. Gleiches gilt für Unterschiede, die durch die Art der Prüfungsvorbereitung bei be-

stimmten Prüfern bewirkt werden können. Die bedeutenderen Unterschiede sind insgesamt bei der Handlungsplanung zu erwarten, weil die Zukunftsperspektive deutlichere Fachunterschiede bewirken müßte.

Die Ausdifferenzierungen des Nachweises der Validität verlangen eine Präzisierung der Auswertung in zwei Richtungen: Zum einen kann die Bestimmung des Ausmaßes der Unterschiede, zum anderen die, allerdings nur begrenzt mögliche, Trennung von Fach- und Prüfereinfluß weiterführen.

Als Methode zur deskriptiven Erfassung des Ausmaßes der von den "Fach-Prüfer-Kombinationen" determinierten Gesamtvarianz der Antworten wird Eta-Quadrat berechnet (Spalte 4 in Tabelle: III.I. - 15). Eta-Quadrat ist eine Schätzung des Varianzanteils der Gesamtvarianz, die auf die Unterschiede in den Bedingungen zurückführbar ist. Zwar ist dieses Maß keine unverzerrte Schätzung (WINER 1971), doch sind die praktischen Verzerrungen so gering, daß das einfach zu berechnende Maß für die hier gestellten Zwecke als ausreichend angesehen werden kann (KERLINGER 1975; NIE et al. 1975).

Im VD ist der Varianzanteil der Unterschiede zwischen den "Fach-Prüfer-Kombinationen" für die Handlungsplanung besonders groß: Er beträgt dort 40 % der Gesamtvarianz, während bei der Beliebtheit nur 17 % der Gesamtvarianz der Antworten durch die entsprechenden Unterschiede bedingt werden. Im HD ist der Varianzanteil bei der Handlungsplanung zwar auch noch, aber doch viel weniger größer als der der Beliebtheit (18 % gegenüber 13 %). Der in beiden Fällen größere Varianzanteil bei der Handlungsplanung entspricht den Erwartungen und bestätigt, daß beide Fragen nach der Motivation Unterschiedliches erfassen, wenn auch damit zu rechnen ist, daß die Handlungsplanung als

überdauernde Perspektive die Beliebtheit teilweise mit determiniert. Der im VD sehr viel größere Varianzanteil der "Fach-Prüfer-Kombinationen" an der Handlungsplanung ist verständlich, da im HD die Fächer enger zusammenhängen und z.B. bestimmte Aspekte eines Faches auch in die an ein anderes angelehnte Berufsperspektive integrierbar sind.

Der Varianzanteil von 16 % für das Selbstkonzept beim HD ist beträchtlich und liegt über den Erwartungen. Hier wird es von der Verursachung durch Prüfer oder Fach abhängen, welcher konkrete Schluß daraus zu ziehen ist. Die Trennung von Fach- und Prüfereinfluß ist auch für die Motivationsfragen wesentlich, da sie letztlich erst die schon gegebene Interpretation absichern kann. Die empirische Kontrolle von Fach und Prüfer als voneinander unabhängige Bedingungen ist, wie in Exkurs II dargestellt und in Kapitel III.I.B. in der Auswertung der Noten praktisch verdeutlicht, nur ansatzweise möglich. Im VD ist der Fächereinfluß in zwei Paaren von "Fach-Prüfer-Kombinationen" zu überprüfen, in denen derselbe Prüfer zwei unterschiedliche Fächer prüft. Der Prüfereinfluß, der dann untersucht werden kann, wenn mindestens zwei verschiedene Prüfer in demselben Prüfungsfach prüfen, kann in drei Fächern an je zwei Prüfern abgemessen werden. Zwar sind an den Vergleichen acht der elf "Fach-Prüfer-Kombinationen" beteiligt; die Selektivität der Vergleiche wird erst daran deutlich, daß aus 55 möglichen Paarungen nur fünf insgesamt einbezogen werden.

Tabelle: III.I. - 16

Mittelwertunterschiede in "Fach-Prüfer-Kombinationen" gleicher Fächer/gleicher Prüfer im VD

	Fach-Prüfer-Kombination I	Fach-Prüfer-Kombination II	arithmet. Mittelwerte FPK I	FPK II	Wahrscheinlichk.zuf. unterschiedl. Mittelw.
PVF-MOT.BEL.					
unterschiedl. Prüfer	Allg.Psychol./12	Allg.Psychol./14	2.19	1.57	<.05
	Pers.Psychol./02	Pers.Psychol./05	2.17	2.07	n.s.
	Philos./01	Philos./03	2.44	2.81	n.s.
unterschiedl. Fächer	Method./02	Pers.Psychol./02	2.50	2.17	n.s.
	Entw.Psychol./03	Pers.Psychol./03	2.05	2.07	n.s.
PVF-MOT.W.A.					
unterschiedl. Prüfer	Allg.Psychol./12	Allg.Psychol./14	2.85	2.57	n.s.
	Pers.Psychol./02	Pers.Psychol./05	2.08	2.17	n.s.
	Philos./01	Philos./03	3.80	3.05	<.01
unterschiedl. Fächer	Method./02	Pers.Psychol./02	2.81	2.08	<.05
	Entw.Psychol./03	Pers.Psychol./03	2.21	2.17	n.s.

Methodisches Vorgehen: Die Varianzen werden mit dem F-Test auf Gleichheit, davon abhängig die Mittelwerte mit dem t-Test für homogene oder heterogene Varianzen geprüft. Ein höherer Mittelwert entspricht minderer Motivation.

Die Ergebnisse sind weder voll den Erwartungen entsprechend, noch widersprechen sie klar. Der erwartete fachbedingte Unterschied ist nur einmal, für die Handlungsplanung, vorhanden. In der Auswertung aber sind die Fächer kaum vertreten, die einen klar "objektiv" verschiedenartigen Zukunftsbezug haben. Der prüferbedingte Unterschied der Handlungsplanung im Fach Philosophie kann auf die Prüferwahl, die dort durch-

gehend möglich ist, zurückgeführt werden. Dies gilt auch für den prüferbedingten Unterschied der Beliebtheit im Fach Allgemeine Psychologie. Die Antworten auf die Motivationsfragen erfahren durch diese Ergebnisse kaum eine zusätzliche Validierung.

Für eine entsprechende Auswertung beim HD stehen vier Zweierpaarungen zur Beurteilung des Prüfereinflusses und nur zwei zur Beurteilung des Fächereinflusses zur Verfügung.

Tabelle: III.I. - 17

Mittelwertunterschiede in "Fach-Prüfer-Kombinationen" gleicher Fächer/gleicher Prüfer im HD

	Fach-Prüfer-Kombination I	Fach-Prüfer-Kombination II	arithmet. Mittelwerte FPK I	FPK II	Wahrscheinlichk.zuf. unterschiedl. Mittelw.
PVF-SK,FÄCHER					
unterschiedl. Prüfer	Sozialpsychol./05	Sozialpsychol./12	5.47	4.54	n.s.
	Päd.Psychol./02	Päd.Psychol./09	5.29	6.21	n.s.
	Klin.Psychol./06	Klin.Psychol./18	6.73	6.40	n.s.
	Angew.Psychol./04	Angew.Psychol./14	4.63	5.53	n.s.
unterschiedl. Fächer	Diagnostik/06	Klin.Psychol./06	5.19	6.73	<.05
	Päd.Psychol./02	Angew.Psychol./02	5.29	3.00	<.05
PVF-MOT,BEL,					
unterschiedl. Prüfer	Sozialpsychol./05	Sozialpsychol./12	2.27	2.31	n.s.
	Päd.Psychol./02	Päd.Psychol./09	2.29	1.93	n.s.
	Klin.Psychol./06	Klin.Psychol./18	1.55	1.60	n.s.
	Angew.Psychol./04	Angew.Psychol./14	2.25	2.33	n.s.
unterschiedl. Fächer	Diagnostik/06	Klin.Psychol./06	2.58	1.55	<.01
	Päd.Psychol./02	Angew.Psychol./02	2.29	2.40	n.s.
PVF-MOT,W,A,					
unterschiedl. Prüfer	Sozialpsychol./05	Sozialpsychol./12	2.47	2.46	n.s.
	Päd.Psychol./02	Päd.Psychol./09	2.86	2.21	n.s.
	Klin.Psychol./06	Klin.Psychol./18	1.55	1.67	n.s.
	Angew.Psychol./04	Angew.Psychol./14	2.88	2.80	n.s.
unterschiedl. Fächer	Diagnostik/06	Klin.Psychol./06	2.27	1.55	<.05
	Päd.Psychol./02	Angew.Psychol./02	2.86	3.00	n.s.

Methodisches Vorgehen: s. Tabelle: III.I. - 16. Höhere Mittelwerte entsprechen einem höheren Selbstkonzept und minderer Motivation.

Die Ergebnisse haben eine klarere Tendenz als die des VD. Offensichtlich trägt der Einfluß der Prüfer wenig zu den Unterschieden bei. Demgegenüber sind bei vier der sechs Paarungen zur Überprüfung der Fachabhängigkeit die Unterschiede signifikant. Die Antworten zu den Motivationsfragen werden dadurch validiert. Für das Selbstkonzept, dessen Antworten zudem in beiden Paarungen Facheinflüsse aufweisen, ist dies dagegen ein erneuter Beweis, daß die Antworten nicht allein vergleichsorientiert, sondern (auch) als Bewertungen kompetenzorientiert gegeben werden. Insbesondere im Fach Klinische Psychologie wird das Kompetenzerleben durch eine intensive praxisbezogene Ausbildung gefördert.

Vergleichend kann zu einer Deutung der Tatsache, daß im HD sich die Fächer sehr viel deutlicher als im VD als die gegenüber Prüfern wesentlichere Einflußquelle herausstellten, auf die Ergebnisse der Befragung von ALVES & LODDERS (1974) verwiesen werden. Die dort festgestellte Dominanz der Aneignung der Fachinhalte im Studium vor dem HD verhindert eine Orientierung der Antworten an den besonderen Bedingungen (wie z.B. den Prüfern) der erwarteten Prüfung. Die demgegenüber im VD deutlich in die Prüfungsvorbereitung verlagerte Aneignung der Inhalte ebnet einem Einfluß der Prüfer den Weg und läßt den Einfluß des Studiums zurücktreten. Dabei darf aber wiederum nicht übersehen werden, daß die Prüflinge des HD durch Selbstselektion vermehrt von der Prüfung unabhängigere Studenten sein können. Zudem ist die unzureichende empirische Basis der Trennung von Fach und Prüfer bei allen Interpretationen in Rechnung zu stellen.

Gewählte Prüfungsordnung und Geschlecht

Die Wahl der alten oder neuen Prüfungsordnung (APO/NPO) kann in Persönlichkeitseigenschaften der Prüflinge begründet sein. Sie stellt zugleich aber auch eine Situationsbedingung dar, indem durch die zeitliche Vorziehung der schriftlichen Hausarbeit bereits vor dem Ausfüllen des PVF eine Rückmeldung erfolgt. Die Wahl der Prüfungsordnung ist nur im Rahmen des VD so verteilt, daß eine Auswertung sinnvoll erscheint. Die beiden Gruppen von Prüflingen weisen weder in den beiden Gesamtkonzepten eigener Fähigkeit noch in den "Über-alle-Fächer" für Selbstkonzept und Motivation errechneten Mittelwerten auffällige und statistisch bedeutsame Mittelwertunterschiede auf.

Wie die Übersicht über die Ergebnisse der Forschung zum Selbstkonzept aufwies, sind zwar nicht einhellig, aber doch häufig Geschlechtsunterschiede feststellbar. Für die Motivationsfragen besteht zu entsprechenden Erwartungen kein Anlaß.

Tabelle: III.I. - 18

Mittelwertunterschiede
zwischen männlichen und weiblichen Prüflingen (nur PVF)

Vordiplom	arithmetische Mittelwerte: männliche Prüflinge	arithmetische Mittelwerte: weibliche Prüflinge	Wahrscheinlichkeit zuf. unterschiedl. Mittelwerte
Gesamtkonzepte			
PVF-SK.GESAMT	5.62	5.25	n.s.
PVF-SK.ZUKUNFT	6.15	6.00	n.s.
Über-alle-Fächer			
PVF-SK.FÄCHER	5.33	5.31	n.s. (?)
PVF-MOT.BEL.	2.46	2.30	n.s. (?)
PVF-MOT.W.A.	2.88	2.92	n.s. (?)

Tabelle: III.I. - 18 (Fortsetzung)

Mittelwertunterschiede
zwischen männlichen und weiblichen Prüflingen (nur PVF)

Hauptdiplom	arithmetische Mittelwerte männliche Prüflinge	weibliche Prüflinge	Wahrscheinlichkeit zuf. unterschiedl. Mittelwerte
Gesamtkonzepte			
PVF-SK.GESAMT	6.44	6.18	n.s.
PVF-SK.ZUKUNFT	6.07	5.33	n.s.
Über-alle-Fächer			
PVF-SK.FÄCHER	5.43	5.68	n.s. (?)
PVF-MOT.BEL.	2.25	2.14	n.s. (?)
PVF-MOT.W.A.	2.35	2.38	n.s. (?)

Methodisches Vorgehen: Die Mittelwerte werden mit t-Test für homogene Varianzen geprüft, da die F-Tests in keinem Fall bedeutsame Unterschiede der Varianzen anzeigen. Höhere Mittelwerte entsprechen höheren Selbstkonzepten und minderer Motivation.

Es ist kein einziger nur annähernd signifikanter Mittelwertunterschied festzustellen. (Auch in der "Weiten Stichprobe" ändert sich das nicht.) Für das Selbstkonzept liegen die Unterschiede zwar für die Gesamtkonzepte in der zu erwartenden Richtung, aber sie sind weit von der Signifikanz entfernt und werden zudem von den einzelfachbezogenen Selbstkonzepten nicht bestätigt.

Da kein sachlicher Grund besteht, einen Einfluß der beiden Bedingungen (Prüfungsordnung, Geschlecht) zu fordern, ist das Fehlen von Unterschieden ein weiterer Beweis der Validität der Antworten.

Zusammenfassung (III.I.E.2.)

Die Analyse der Situations- und Persönlichkeits-

bedingungen des Antwortverhaltens dient der Validierung der Antworten. Die Unterschiede zwischen den beiden Prüfungsarten (VD und HD) können nicht allein Ergebnis der Einwirkung der Selbstselektion beim HD sein. Sie lassen vermuten, daß die Antworten für die Gesamtkonzepte eigener Leistungsfähigkeit kompetenzorientiert gegeben werden. Die "höhere" Motivation der Prüflinge des HD entspricht der objektiven Situation und stellt eine Konstruktvalidierung der Fragen dar; die Beliebtheit scheint von der Handlungsplanung überformtzu werden.

Mittelwertunterschiede zwischen den "Fach-Prüfer-Kombinationen" sind im VD besonders für die Handlungsplanung, auch für die Beliebtheit, nicht aber für das fachbezogene Selbstkonzept festzustellen. Im HD dagegen sind für alle drei Fragen die Mittelwerte unterschiedlich. Die deutlichsten Unterschiede sind wiederum in der Handlungsplanung vorhanden, doch längst nicht in dem Ausmaß, wie das beim VD der Fall ist. Die Ergebnisse stellen überwiegend Validierungen der Fragen dar: Im VD wird die relative Unabhängigkeit der beiden Motivationsfragen verdeutlicht; die fragwürdigen Unterschiede für das Selbstkonzept im HD können auf die schon erwähnte kompetenzorientierte Abgabe der Antwort zurückgehen. Der Versuch einer Trennung des Fach- oder Prüfereinflusses zeigt, daß im VD die Unterschiede auch prüferbedingt sein können. Das schränkt die Validität der Antworten etwas ein, kann aber auf die objektiv wesentlich größere Bedeutung der Prüfung zurückgeführt werden. Die Unterschiede des HD sind, soweit sich das überprüfen läßt, eindeutig den Fächern zuzuordnen.

Gewählte Prüfungsordnung und Geschlecht der Prüflinge ergeben keine bedeutsamen Unterschiede in den Mittelwerten der Antwortverteilungen. Das ist als ein Hinweis auf die Rationalität der Antworten zu verstehen.

III.I.E.3.: Kontingenzen der thematischen Fragen

Die Zusammenhänge zwischen den thematischen Fragen umfassen zwei unterschiedliche empirische Zugriffe. 1. Die Binnenbeziehungen innerhalb der einzelfachbezogenen Fragen einer thematischen Frage sind ein Maß der Situationsspezifität des Antwortverhaltens und geben weitere Aufschlüsse über die Validität des Antwortverhaltens. 2. Die Beziehungen zwischen den thematischen Fragen gehen das Problem der auf der theoretischen Ebene bereits erörterten prozessualen und strukturellen Selbständigkeit an. Sie sind ein Beitrag zur differentiellen Validität der einzelnen thematischen Fragen.

Die Berechnung der Kontingenzen erfolgt durchweg über Produkt-Moment-Korrelationen, da die Verteilungsvoraussetzungen für deren Signifikanzprüfung erfüllt sind. Probleme der Signifikanzprüfung durch Berechnung mittlerer Korrelationen und Verwendung partiell abhängiger Daten werden an Ort und Stelle diskutiert.

Kontingenzen innerhalb der Fragen

Die über die Prüflinge hinweg berechneten Korrelationen der Antworten zu derselben Frage in zwei verschiedenen Einzelfächern sind ein Maß der situativen Differenziertheit der Antworten, wenn die Korrelation über alle Kombinationen von je zwei Fächern hinweg berechnet und gemittelt wird. Besteht ein hoher gemittelter Zusammenhang zwischen den Antworten zu den einzelnen Fächern, dann kann man die Antwort eher auf situationsunabhängige Eigenarten der Person zurückführen. Bei niedrigem mittleren Zusammenhang können die einzelfachbezogenen Antworten durchgehend voneinander unabhängig sein oder, von den strukturellen Beziehungen zwischen den Fächern bedingt, unterschiedliche Zusammenhänge aufweisen. Im letzteren Fall wür-

den die Korrelationen stark streuen. Beidemal liegt eine situationsspezifische Beantwortung vor; im letzten Fall ist sie positiv bewiesen; bei durchgehend geringen Korrelationen könnte das auch auf die geringe Zuverlässigkeit der Antworten zurückgeführt werden.

Die nachgewiesenen Mittelwertunterschiede zwischen den einzelnen "Fach-Prüfer-Kombinationen" sind nur ein indirekter Hinweis auf die Situationsspezifität der individuellen Antworten, da durchaus die Möglichkeit besteht, daß die in den Mittelwerten verschobenen Antwortverteilungen sich in der Binnenstruktur, die sich als Profil der individuellen Einschätzungen über die Fächer hinweg darstellt, weitgehend gleichen. Die mittleren Korrelationen zwischen allen Paaren von einzelfachbezogenen Antworten sind ein Maß für die interindividuell unterschiedliche Gestalt der Antwortprofile.

Die Ergebnisse lassen im Sinne der normativen Rahmenvorstellungen unter den Leitideen von Individualisierung bezüglich der Motivation und Rationalität bezüglich des Selbstkonzepts geringe mittlere Korrelationen und eine bedeutsame Streuung der Korrelationen erwarten. Das baut auf der Annahme auf, daß sowohl die Handlungsplanungen als auch die Fähigkeiten in den verschiedenen Fächern individuell unterschiedlich sind, zugleich aber einzelne Fächer eine bedeutsame inhaltliche Überschneidung aufweisen, die in einzelnen höheren Korrelationen ihren Ausdruck finden sollte (so gehören Klinische Psychologie und Diagnostik insgesamt eng zusammen).

Tabelle III.I. - 19

Streubereich und Mittelwerte der Korrelationen zwischen den einzelfachbezogenen Antworten im PVF

	PVF-SK.FÄCHER	PVF-MOT.BEL.	PVF-MOT.W.A.
Vordiplom			
mittlere Korrelation	+.38	+.07	+.05
Streubereich der einzelnen Korrelationen	-.15 bis +.64	-.13 bis +.43	-.37 bis +.56
Hauptdiplom			
mittlere Korrelation	+.32	-.05	-.11
Streubereich der einzelnen Korrelationen	+.16 bis +.56	-.29 bis +.37	-.65 bis +.55

Methodisches Vorgehen: Die mittleren Korrelationen werden als arithmetische Mittelwerte in den 21 möglichen Fächerpaarungen des VD und den 10 möglichen Fächerpaarungen des HD berechnet und an der Personenzahl gewichtet gemittelt (im VD sind in den Nebenfächern weniger Prüflinge beteiligt, weil ihnen Prüfungen anerkannt werden). Eine z-Transformation ist nicht notwendig, weil besonders hohe Korrelationen fehlen. Eine Signifikanzbestimmung für die mittleren Korrelationen ist problematisch; zumindest gelten die Signifikanzgrenzen der einzelnen Korrelationen. Sie betragen im VD: ab $r = .30$: $p < .05$; ab $r = .39$: $p < .01$. Sie betragen im HD: ab $r = .37$: $p < .05$; ab $r = .47$: $p < .01$. Da die mittleren Korrelationen zumindest die Zahl der Freiheitsgrade haben dürften, die sich aus der Zahl der Prüflinge ergibt, ist die mittlere Korrelation des Selbstkonzepts im VD signifikant.

Die Zusammenhänge beim Selbstkonzept sind bis auf eine einzige Fächerpaarung positiv und erreichen

in einzelnen Paaren einen engen Zusammenhang: r = .64 im VD und r = .56 im HD als höchste Werte, beide sind signifikant mit $p < .01$. Zwar bleibt bei der mittleren Korrelation im VD und HD von r = .38 bzw. r = .32, wenn man die Zuverlässigkeit der Antworten wesentlich höher einschätzt (s. Exkurs I), noch der weitaus größere Varianzanteil für die situationsspezifische Varianz der Antworten, doch ist die personale Konstanz eines wesentlichen Anteils der einzelfachbezogenen Selbstkonzepte nicht zu übersehen. Die Bedeutung eines personal stabilen Selbstkonzepts mag u.a. darauf zurückzuführen sein, daß die Ausbildung selbst gar nicht so streng nach den einzelnen Fächern geschieden ist, wie das die Prüfung mit ihren Fächergliederungen vorgibt. Dieser Vermutung widerspricht aber die im VD höher als im HD ausfallende mittlere Interkorrelation. Im VD hat die Vorbereitung, die fachspezifisch organisiert ist, einen wesentlich größeren Anteil am Erwerb kognitiver Strukturbildungen (ALVES & LODDERS 1974); dort müßten also solche fächerspezifischen Erfahrungen vorliegen.

Eine alternative Erklärung ist darin zu suchen, daß die Zusammenhänge unter Bedingungen enger werden, in denen fachspezifische Kompetenzerlebnisse fehlen. Dem würde die höhere mittlere Korrelation des VD entsprechen.

Die mittleren Korrelationen für die beiden Motivationsfragen dagegen entsprechen den Erwartungen. Sie weisen weder im HD noch im VD eine bedeutsame Höhe auf. Die Motivationsmaße erfassen einen Person/Umwelt-Bezug, der individuell unterschiedlich ist. Wie aus den Extremwerten einzelner Korrelationen hervorgeht, sind diese Unterschiede in dem individuellen Bezug von Gegenstand und Prüfling begründet. Der Gegenstand der Motivation kann, wie auf Konstruktebene vermutet, über einzelne Fächer hinausgehend Einheiten

bilden. Dabei zeigt sowohl im VD als auch im HD der größere Streubereich der Korrelationen bei der Handlungsplanung eine prägnantere Gegenstandsstrukturierung als bei der Beliebtheit an; ein weiterer Beitrag zur differentiellen Validität der beiden Fragen. Die Einebnung der Prägnanz der Gegenstandsstruktur bei der Beliebtheit könnte in der mehr auf die augenblickliche Situation bezogenen emotionalen Bewältigung der Prüfungsvorbereitung begründet sein.

Die Ergebnisse machen deutlich, daß die statistische Problematik der Auswertung "Über-alle-Fächer" und Verrechnung der Antworten zu einzelnen Fächern als voneinander unabhängige Daten für die Motivationsfragen nicht ganz so schwerwiegend ist, da dort Abhängigkeiten nur partiell bestehen, während für das Selbstkonzept infolge der durchgängigen positiven Abhängigkeiten mehr Bedenken anzubringen sind.

Eine weitere Klärung der Binnenstruktur der fachbezogenen Selbstkonzepte geht aus ihrer Beziehung zu den Gesamtkonzepten hervor. Zusätzlich sind die Zusammenhänge zwischen den Gesamtkonzepten wiedergegeben.

Tabelle: III.I. - 20

Korrelationen zwischen fachbezogenen Selbstkonzepten und Gesamtkonzepten im PVF

	Einzelfächer							
	Ausdr./ Method.	Allg. Psychol.	Entw. Psychol.	Pers. Psychol.	Physiol.	Wahl- fach I	Wahl- fach II	PVF- SK.GESAMT
Vordiplom								
PVF-SK.GESAMT	.56++	.47++	.61++	.54++	.59++	.34+	.41+	
PVF-SK.ZUKUNFT	.31+	.38+	.38+	.19	.31	.09	.51++	.51++

Tabelle: III.I. - 20 (Fortsetzung)

Korrelationen zwischen fachbezogenen Selbstkonzepten und Gesamtkonzepten im PVF

	Einzelfächer					
	Diagnost.	Sozial-psychol.	Päd. Psychol.	Klin. Psychol.	Angew. Psychol.	PVF-SK.GESAMT
Hauptdiplom						
PVF-SK.GESAMT	.60++	.67++	.22	.66++	.62++	
PVF-SK.ZUKUNFT	.51++	.30	.10	.36	.69++	.71++

Methodisches Vorgehen: Die Signifikanzen der Korrelationen werden bei $p < .05$ durch "+", bei $p < .01$ durch "++" symbolisiert. Durch den Ausfall einzelner Antworten liegen die Signifikanzgrenzen im VD nicht für alle Fächer gleich hoch.

Zu den Zusammenhängen zwischen PVF-SK.GESAMT und den fächerbezogenen Selbstkonzepten muß einschränkend angemerkt werden, daß die Anordnung der Fragen im Fragebogen zu einer Vergrößerung der wirklichen Zusammenhänge beitragen kann, indem die Frage nach dem Gesamtkonzept den Fragen nach den Einzelkonzepten nachgeordnet ist und die Beantwortung rein optisch als Durchschnittsbildung vorgenommen werden kann. (Hinzu kommt in einigen Fällen die entsprechende Einsetzung bei fehlender Antwort im PVF-SK.GESAMT; s. Exkurs I.)

Der Zusammenhang zwischen dem gegenwartsbezogenen Gesamtkonzept eigener Fähigkeit und dem zukunftsbezogenen ist im VD nicht so eng, als daß nicht eine relative Selbständigkeit beider anzunehmen ist. Demgegenüber läßt der Zusammenhang im HD weniger eine funktionale Differenzierung erwarten. Die Zunahme des Zusammenhangs ist Ausdruck der größeren Identität von

Studium vor der Prüfung und Tätigkeit nach der Prüfung im HD. Aber auch im HD ist wie im VD das Muster der Beziehungen zwischen fachbezogenen Selbstkonzepten und zukunftsbezogenem Gesamtkonzept ein anderes als das der Beziehungen zum gegenwärtigen Gesamtkonzept, was doch für eine funktionale Differenzierung beider Gesamtkonzepte im HD spricht.

Die einzelnen Korrelationen sind ohne genaue Analyse der einzelnen Fachbedingungen kaum differenziert auszuwerten. Im VD fällt die relativ gleich enge Beziehung der Selbstkonzepte in den zentralen Fächern (unter Einschluß von Physiologie) zum Gesamtkonzept auf. Sie legt die Annahme einer gleichmäßigen Überformung durch das Gesamtkonzept nahe (die auf fehlende differenzierte Erfahrung zurückgehen kann; s. oben). Diese Annahme ist auf das HD nicht ohne Bedenken zu übertragen. Das Selbstkonzept in Pädagogischer Psychologie hat nur einen geringen Zusammenhang mit dem Gesamtkonzept. Sieht man dies auf dem Hintergrund des Fehlens jeder systematischen Ausbildung in diesem Fach zur damaligen Zeit, dann ist gleichzeitig die Interpretation der Überformung durch das Gesamtkonzept in Frage gestellt, denn dieses müßte gerade in einer solchen Bedingung auf ein wenig von anderen Kompetenzerlebnissen bedingtes Selbstkonzept durchschlagen. Hier können nur differenziertere Untersuchungen weitere Klarheit schaffen.

Kontingenzen zwischen den Fragen

Die Feststellung funktionaler Selbständigkeit (differentieller Validität) der Antworten wird ergänzt durch eine Analyse der direkten Zusammenhänge zwischen den thematischen Fragen. Weitere Hinweise auf mögliche kausale Verknüpfungen sind davon aber nicht zu erwarten. Insofern ist auch das in Teil D angesprochene

Problem möglicher "Einmischung" der Motivation in Antworten zum Selbstkonzept hier kaum empirisch voll zu klären.

Tabelle: III.I. - 21

Korrelationen zwischen den thematischen Fragen im PVF

	Fach-Prüfer-Kombinationen			Über-alle-Fächer	Mittelwerte der Fach-Prüfer-Komb.	
	mittl. r	von bis	Zahl der sign. r's mit $p<.05$	r	r	rho
Vordiplom						
PVF-SK.FÄCHER - PVF-MOT.BEL.	-.34	-.12 -.70	4	-.42	-.52	-.59
PVF-SK.FÄCHER - PVF-MOT.W.A.	-.15	.13 -.55	2	-.20	-.38	-.38
PVF-MOT.BEL. - PVF-MOT.W.A.	.52	.29 .74	8	.55	.64+	.65
Hauptdiplom						
PVF-SK.FÄCHER - PVF-MOT.BEL.	-.41	-.01 -.65	2	-.43	-.73+	-.72
PVF-SK.FÄCHER - PVF-MOT.W.A.	-.32	.29 -.93	2	-.45	-.78++	-.76
PVF-MOT.BEL. - PVF-MOT.W.A.	.53	.15 .76	5	.58	.80++	.56

VD: df = 40 $r > .30$ bei $p < .05$
$r > .39$ bei $p < .01$

HD: df = 27 $r > .37$ bei $p < .05$
$r > .47$ bei $p < .01$

Methodisches Vorgehen: Die Ebenen der Auswertung entsprechen den in Exkurs II dargestellten. In den "Fach-Prüfer-Kombinationen" wird der durch individuelle Unterschiede vermittelte Zusammenhang deutlich; auf der Ebene "Über-alle-Fächer" kommen Fach und Prüfer hinzu. Die Korrelation der Mittelwerte erfaßt allein die durch Fach und Prüfer bedingten Unterschiede (in allem gelten die Einschränkungen von Exkurs II).

Die Mittelung der Korrelationen in Spalte 1 stellt das arithmetische Mittel der nach Freiheitsgraden gewichteten z-transformierten Produkt-Moment-Korrelationen in den einzelnen "Fach-Prüfer-Kombinationen" dar. Die Korrelationen in den "Fach-Prüfer-Kombinationen" streuen stark, was u.a. durch die teilweise geringe Personenzahl in bestimmten Kombinationen (besonders im HD) bedingt wird. Die Signifikanzprüfung der gemittelten Korrelationen, die partiell voneinander abhängig sind, da sich die Prüflinge der "Fach-Prüfer-Kombinationen" überschneiden, kann zumindest die Freiheitsgrade für die Zahl der Prüflinge voraussetzen. Die Signifikanzgrenzen dafür sind unter der Tabelle angeführt.

In den letzten beiden Spalten sind die über die 11 (VD) bzw. 10 (HD) Kombinationen berechneten Korrelationen der Mittelwerte wiedergegeben. Zusätzlich zur Produkt-Moment-Korrelation werden noch entsprechende Rangkorrelationskoeffizienten angegeben, die den Effekt von "Ausreißern" auf die Produkt-Moment-Korrelationen deutlich machen sollen. Wie zu sehen ist, ist nur in einem Fall eine größere Abweichung zwischen beiden Zusammenhangschätzungen gegeben, so daß die Auswertung über die Produkt-Moment-Korrelationen vorgenommen wird. Dort sind signifikante Korrelationen durch "+" für $p < .05$ und "++" für $p < .01$ gekennzeichnet.

Rein formal gesehen wird deutlich, daß die Zusammenhänge im wesentlichen schon in den "Fach-Prüfer-Kombinationen" bestehen, also individuell bedingt sind. Es handelt sich dabei aber nicht um individuelle Konstanten im Sinne von personengebundenen Zusammenhängen, sondern um situationsspezifisch differenzierte individuelle Beziehungen. Die durch Fach- und Prüferunterschiede bedingten zusätzlichen Zusammenhänge

tragen nicht wesentlich zu einer Veränderung der Korrelationen bei, sie sind auch weder in Richtung noch im Umfang deutlich abweichend; zudem ist, wie gezeigt wurde, mit Ausnahme der Handlungsplanung im VD weniger als 20 % der jeweiligen Antwortvarianz durch Fach und Prüfer determiniert.

Der Zusammenhang zwischen der Beliebtheit und der Handlungsplanung umfaßt unter Berücksichtigung aller gleichgerichteten Kovarianzbeziehungen ("Über-alle-Fächer") in beiden Prüfungsarten nur um 30 % der Varianz der Antworten in beiden Fragen. Eine relative Selbständigkeit der beiden Aspekte der Motivation ist deshalb gegeben. Die Beliebtheit kann ihre Funktion als Kontrollfrage gegenüber der Handlungsplanung erfüllen.

Der Zusammenhang des Selbstkonzepts mit den beiden Motivationsfragen soll eine Antwort auf die mögliche motivationale Verzerrung des Selbstkonzepts im PVF geben. Die Korrelationen sind auf den Ebenen individueller Antworten insgesamt nicht besonders hoch. Im VD spricht der engere Zusammenhang zwischen Selbstkonzept und Beliebtheit für die Möglichkeit einer Einmischung der Motivation; im HD ist dieser Unterschied zwischen Beliebtheit und Handlungsplanung aber nicht festzustellen. Dort fallen die bedeutsamen Korrelationen der Mittelwerte von Selbstkonzept und Motivationsfragen ins Auge, die die Vermutung bestätigen, daß das Selbstkonzept zumindest im HD sich an der erlebten Kompetenz orientiert.

Zusammenfassung (III.I.E.3.)

Die Zusammenhänge innerhalb der Antworten zu den einzelfachbezogenen Fragen machen deutlich, daß die Antworten zum Selbstkonzept eine bedeutsame individuell konstante Komponente aufweisen, während die Antworten

zu den Motivationsfragen im Mittel zwischen den Einzelfächern nur gering korrelieren. Der dort große Streubereich der einzelnen Korrelationen zwischen den Fächern ist auf eine Gegenstandsstruktur der Motivation zurückzuführen, die über die Fächergliederungen hinausgeht. Diese Struktur ist bei der Handlungsplanung wesentlich prägnanter ausgeprägt als bei der Beliebtheit. Das verweist auf die differentielle Validität der beiden Motivationsfragen. Die Zusammenhänge zwischen den Antworten in den Einzelfächern lassen eine statistische Auswertung "Über-alle-Fächer" hinweg insbesondere bei den Motivationsfragen als nicht ganz so problematisch erscheinen.

Die Zusammenhänge zwischen den thematischen Fragen zum Selbstkonzept und zur Motivation zeigen auf formaler Ebene, daß die Kovarianz der einzelfachbezogenen Antworten im wesentlichen auf der Ebene der situationsspezifischen individuellen Beantwortung determiniert wird, während Fach und Prüfer nur wenig zu den Zusammenhängen beitragen. Der Zusammenhang zwischen den beiden Motivationsfragen ist nicht so eng, als daß nicht eine funktionale Unabhängigkeit von Handlungsplanung und Beliebtheit gesichert wäre. Die Korrelationen von Selbstkonzept und beiden Motivationsfragen bleiben auf der Ebene individueller Antworten unter einem Varianzanteil von 20 %. Im VD ist eine Einmischung der Motivation in das Selbstkonzept nicht auszuschließen, während im HD der Zusammenhang mehr durch die Kompetenzorientierung des Selbstkonzepts vermittelt sein kann.

III.II.: Wirkungsanalyse

Die Wirkungsanalyse versucht in den beiden Teilen dieses Kapitels, im Teil A mit den theoretischen und empirischen Erfahrungen, die sich in der Psychologie angesammelt haben, im Teil B mit der Auswertung der Daten der empirischen Untersuchung, Möglichkeit, Ausmaß und Verfälschung der direkten Wirkung der Prüfung, insbesondere der im Verlauf der Prüfung erteilten Noten, auf Selbstkonzept und Motivation der Prüflinge zu ermitteln. Die in Kapitel II aufgestellten Normwerte sind bei diesem Vorgehen leitend, wenngleich durch sie nicht jeder Schritt des Vorgehens abgedeckt ist, weil sowohl im berichtenden Teil als auch im Teil der empirischen Auswertung sich vom Gegenstand her weitere Aspekte aufdrängen.

III.II.A.: Theoretische Klärung möglicher Wirkungen

Übersicht: Wie schon bei der Voraussetzungsanalyse fallen unter den theoretischen Teil neben eigentlich theoretischen Ansätzen auch eklektische begriffliche Erörterungen sowie "Übertragungen" der Ergebnisse von Untersuchungen, die selbst keinem expliziten theoretischen Kontext entstammen.

Der Teil 1 der theoretischen Klärung gilt dem Selbstkonzept eigener Fähigkeit. Durch Heranziehung sozialpsychologischer Theorien werden Erkenntnisse über wahrscheinliche Wirkungen gewonnen. Das ergänzt die Aufarbeitung von Erfahrungen mit Rückmeldungen im Bereich der diagnostischen Psychologie und von einigen wenigen Untersuchungen, die die Rückmeldung mit Noten erforscht haben. Zusätzlich wird ansatzweise auf der Grundlage dieser Erkenntnisse eine Klärung der konkreten Situation, in der die Wirkung der Prüfung stattfindet, versucht.

Der entsprechende Teil 2 zu den Grundlagen der Wirkung auf die Motivation kann wiederum knapper ausfallen, da die vorliegenden Erkenntnisse nur unzureichend situationsspezifisch interpretierbar sind. Das trifft sich mit der normativ begründeten größeren Bedeutung des Selbstkonzepts eigener Fähigkeit.

Vorab ist eine kurze Präzisierung der Wirkungsanalyse notwendig, die zugleich eine Eingrenzung des Begriffes der Rückmeldung beinhaltet.

Eingrenzung der Wirkungsanalyse

Der berichtende (theoretische) Teil dieses Kapitels ist von der empirischen Auswertung der Untersuchung relativ abgelöst und selbständig (s. einleitende Bemerkungen zu Kapitel III.). Das macht ein unterschiedliches Vorgehen des theoretischen und empirischen Teils sinnvoll: Im empirischen Teil wird im wesentlichen der globale Zusammenhang zwischen den unabhängigen Variablen, besonders den Noten, und den abhängigen Variablen, Selbstkonzept und Motivation, untersucht. Einzelne Prozesse, die in diesen Wirkungszusammenhang vermittelnd eingreifen können, werden nicht beachtet. Im berichtenden Teil dagegen nehmen solche Prozesse einen wichtigen Stellenwert ein. Sie erst machen es möglich, in einer konkreten Situation die Chancen abzusehen, die für eine Wirkung im Sinne der Normwerte sprechen.

Die gewonnene Erkenntnis soll für die vorliegende Situation auch solche Aspekte herausstellen, die einer eigenständigen empirischen Überprüfung nicht zugänglich sind. Dabei besteht allerdings ein grundlegendes Problem: Die Übertragung kann nur dann Gültigkeit beanspruchen, wenn die situativen (und personalen) Bedingungen eine Übertragbarkeit von Erkenntnissen zulassen. Diese Bedingungen aber sind nicht genü-

gend bekannt. Eine zufriedenstellende Lösung für dieses Problem wäre allein in einem mehrstufigen Vorgehen zu finden, das zusätzlich die Übertragbarkeit selbst empirisch absichert. Solange das nicht möglich ist, bleibt nur der Ausweg, den Rahmen für Übertragungen möglichst weit zu stecken. In dieser Absicht werden im theoretischen Teil aus unterschiedlichen Gegenstandsbereichen der Psychologie und auf unterschiedlichem begrifflichen und empirischen Niveau Ergebnisse zusammengetragen. Das begleitet die Erwartung, aus der Fülle verschiedenartiger Ansätze zu einem zusammenfassenden Urteil zu kommen. Dabei geht dann jedoch die Situationsspezifität verloren. Deshalb wird abschließend eine Integration der Ergebnisse versucht, die eine vorläufige, empirisch nicht bestätigte Deutung der Besonderheit der Situation einbezieht.

Grenzen für die Übertragbarkeit von Erkenntnissen sind dennoch vorgreifend zu ziehen. Sie sind dort besonders wichtig, wo sich hinter gleichen Begriffen unterschiedliche Prozesse oder Strukturen verbergen. Eine zentrale Bedeutung hat für die Übertragung die Präzisierung des Vorgangs der Rückmeldung, als der die Wirkung der Noten betrachtet werden kann. Auf Grund der begrifflichen Bestimmung liegt es nahe, sich mit den Ergebnissen von empirischen Untersuchungen zu einem Prozeß zu beschäftigen, der als Ergebnisrückmeldung (knowledge of results, LOCKE et al. 1968; knowledge of performance, AMMONS 1956) bezeichnet wird. Bei genauerer Hinsicht aber stellt sich heraus, daß die in den zugehörigen Experimenten dem Begriff zuzuordnenden experimentellen Prozeduren kaum etwas mit den Prozessen, die in Prüfungen stattfinden, gemein haben. Ein großer Teil der Experimente, aus denen AMMONS (1956) seine Generalisierungen über die Rückmeldung herleitet, besteht aus einfachen motorischen

Aufgaben. Es handelt sich dabei wie auch bei den übrigen Experimenten um kurzzeitig erstreckte Lernprozesse, in denen die Rückmeldung sich auf die konkreten Konsequenzen eines ebenso konkreten, relativ einfachen Lernprozesses bezieht. SASSENRATH et al. (1968), die zum selben Thema Versuche mit einzelnen Wissenstestitems anstellen, verwenden in diesem Zusammenhang den Begriff der informativen Rückmeldung. Doch liegt der Unterschied zwischen diesen Experimenten und der Rückmeldung im Rahmen von Prüfungen über Noten nicht darin, daß letztere dann als normative Rückmeldung zu bezeichnen wäre (in Anlehnung an die Unterscheidung von informativem und normativem sozialen Einfluß bei TEDESCHI (1972), s. unten). Die Rückmeldung über Noten beinhaltet zweifellos eine wertende Komponente; aber das macht nicht den bestimmenden Unterschied zu den Untersuchungen über unmittelbare Rückmeldung aus, sondern die Komplexität der Prozesse ist unterschiedlich. Während in den Experimenten der einzelne Lernakt eine Rückmeldung erhält, betrifft die Rückmeldung in der Prüfung eine ganze Lernsequenz und damit die übergeordnete Steuerung einzelner Lernakte. Dementsprechend müssen die Prozesse, in denen die Rückmeldung verarbeitet wird, unterschiedlich sein: Im Falle einzelner Lernakte ist eine unmittelbare Verhaltensänderung möglich; im Falle von Lernsequenzen findet die Änderung erst einmal auf der Ebene übergeordneter Steuerungsprozesse kognitiv-motivationaler Art statt (wie Selbstkonzepten, Handlungsplanung usw.). Die Ergebnisse über die elementare Form der Rückmeldung finden deshalb keine Beachtung.

III.II.A.1.: Wirkungen auf das Selbstkonzept eigener Fähigkeit

Konstitution der Selbstkonzepte in der Sozialisation

Vom Ausgangspunkt der theoretischen Verwendung in der Sozialpsychologie her gesehen werden Selbstkonzepte als grundsätzlich durch Sozialisationsprozesse herausgebildet verstanden (GERGEN 1971). Spuren sozialbedingter Genese sind in der Ähnlichkeit des Selbstkonzepts eigener Fähigkeit von Kindern mit dem ihrer Eltern (BILBY et al. 1972) oder der für gesellschaftliche Gruppen unterschiedlichen Ausprägung von Selbstkonzepten (WYLIE 1963) zu finden. An dieser Stelle ist mit der Untersuchung von WYLIE auch auf die in Kapitel III.I.C. berichteten Unterschiede zwischen den Geschlechtern usw. zu verweisen.

Verschiedenartige Quellen werden dabei nicht einfach passiv aufgenommen, sondern aktiv verarbeitet. Das hat zur Folge, daß im Selbstkonzept nicht so sehr der Einfluß bestimmter einzelner Quellen nachweisbar ist, sondern mehr der integrierte Einfluß verschiedener (BROOKOVER & THOMAS 1964).

Der Sozialisationsprozeß von Selbstkonzepten eigener Leistungsfähigkeit geht unter maßgeblicher Beteiligung der schulischen Bildungsprozesse vor sich: Zwischen der 6. und 12. Klasse stellen BAILEY & GIBBY (1971) eine deutliche Annäherung des Selbstkonzepts an die Ergebnisse von Intelligenztests fest; das Selbstkonzept eigener intellektueller Fähigkeit wird realistischer. Einflüsse sind auch im tertiären Bereich zu beobachten (KEEFER 1971). Dabei darf aber nicht übersehen werden, daß gerade die schulische Sozialisation des Selbstkonzepts eigener Leistungsfähigkeit nicht als ein passiver Prägevorgang verstanden werden darf. Das Selbstkonzept der Fähigkeit beeinflußt den Einsatz des Schülers und bewirkt auf diese Weise eine Umkehrung

der Kausalität, indem das Selbstkonzept für die zunehmende Übereinstimmung mit der Leistung verantwortlich wird. Solche wechselseitigen Abhängigkeiten sind z.B. beim Zueinander von Selbstkonzept und Underachievement anzunehmen (s. oben, Kapitel III.I.C.; DAVIDSON & LANG (1960)).

Das theoretische Modell der Herausbildung der Selbstkonzepte (eigener Fähigkeit) im Sozialisationsprozeß sichert im Zusammenhang mit empirischen Nachweisen die Möglichkeit, daß Selbstkonzepte Erfahrungen, insbesondere solchen, die sozial vermittelt sind, zugänglich sind. Ungeklärt ist, inwieweit gegenständlich spezifische Selbstkonzepte, z.B. der Leistungsfähigkeit in einem bestimmten Fach, durch Ausdifferenzierung allgemeinerer entstehen, z.B. eines Konzepts eigener Begabung (MEYER 1973), oder eine relativ selbständige Genese aufweisen. Deshalb erfolgt ohne Zwischenschritt der Übergang zur Verarbeitung rein situativ bedingter Erfahrungen in Selbstkonzepten eigener Fähigkeit.

Situativer Einfluß auf Selbstkonzepte

Diese Themenstellung entspricht dem Ansatz der Untersuchung. Die Wirkung der Prüfung (und der Noten) auf das Selbstkonzept des Prüflings ist ein situativ gebundener, einmaliger Prozeß, der sich deutlich von den zeitlich langerstreckten und strukturell vielgestaltigen Einflüssen in Sozialisationsprozessen abhebt. Die letzteren wiederum können neben anderen auch solche einmaligen Einflüsse beinhalten. Vorliegende Erkenntnisse zum Einfluß von bestimmten Situationen grenzen die Wirkungen auf Änderungen ein. Prinzipiell ist auch ein Einfluß denkbar, der eine Bestätigung eines vorhandenen Selbstkonzepts beinhaltet und nicht zu einer Änderung Veranlassung gibt. Ein solcher Ein-

fluß könnte sich in der Festigkeit, z.B. in der Form subjektiver Sicherheit des entsprechenden Selbstkonzepts, auswirken. Das bleibt hier ausgeklammert (s. unten).

In der Prüfung ist die Situation dadurch gekennzeichnet, daß der Prüfer Bewertungen äußert; dementsprechend werden hier Erkenntnisse ausgewertet, die den Einfluß von - meist direkt übermittelten - Fremdbewertungen auf Selbstkonzepte betreffen.

Die an eine bestimmte Situation gebundene Wirkung einer kommunikativ übermittelten Fremdbewertung auf das Selbstkonzept ist in dreierlei Hinsicht begrifflich und empirisch näher zu bestimmen: Es sind die allgemein sozialpsychologischen Grundlagen eines solchen Prozesses sozialen Einflusses zu erörtern, die die Bedingung der Möglichkeit einer Änderung darstellen. Daran schließt sich ein zweiter Problemkreis an, der sich mit den situativen Bedingungen eines solchen Einflußprozesses beschäftigt. Neben den besonderen situativen Bedingungen sind auch noch besondere personale Bedingungen anzusetzen, die für bestimmte Personen unterschiedliche Wirkungen erwarten lassen und den allgemeinen Einfluß verfälschen können.

Die <u>(sozial)psychologischen Grundlagen</u> einer situativen Beeinflussung des Selbstkonzepts gehen davon aus, daß das Selbstkonzept in jeder gegebenen Situation in adaptiver Weise dem eigenen Handeln und den Ergebnissen eigenen Handelns Bedeutung verleiht (BENJAMINS 1950). Dies erfolgt dabei nicht unabhängig von bereits gebildeten Selbstkonzepten. Das jeweils vorhandene Selbstkonzept beeinflußt nach LOWE (1961) die Aufnahme weiterer selbstkonzeptbezogener Erfahrungen. Es weist eine Beharrungstendenz auf, indem Personen gemäß seiner konkreten Ausprägung aktiv bestätigende Vergleichsinformationen suchen (FESTINGER et al. 1954) oder durch

Fehlwahrnehmung (NEUBAUER 1976) abweichende Erfahrungen abwehren. Das Ausmaß der Beharrungstendenz wird von dem Ausmaß der Festigkeit des gebildeten Selbstkonzepts abhängen. So ist nach sensorischer Deprivation das Selbstkonzept weniger gegen Änderungen stabil (JONES & SCHNEIDER 1968).

Zwei konkurrierende Theorien versuchen, die Reaktionen auf Fremdbewertungen, die von Selbstbewertungen abweichen, zu erklären: die Dissonanztheorie und die Theorie des Selbstwertgefühls (self-esteem theory). Die Dissonanztheorie ist die weitaus allgemeinere der beiden Theorien, die nicht nur auf selbstbezogene Einstellungen, als die Selbstkonzepte aufgefaßt werden können, anwendbar ist. Sie geht als Kongruenztheorie davon aus, daß Personen (motivational zu begründende (?); FELDMAN 1966) Tendenzen zur Auflösung von einander widersprechenden Kognitionen zeigen (FESTINGER 1957). Fremdbewertungen, die vom eigenen Selbstkonzept abweichen, stellen u.a. solche widersprüchlichen Kognitionen dar. (Das Problem einer genaueren psychologischen Bestimmung solcher Widersprüche kann an dieser Stelle außer acht gelassen werden; s. auch IRLE 1975.) Der Dissonanz erfahrenden Person stehen nun eine Reihe von Möglichkeiten zur Verfügung, die Inkongruenzen zwischen den Kognitionen zu verringern. Die konkrete "Wahl" bestimmter Möglichkeiten ist für das Problem der Änderung des Selbstkonzepts ausschlaggebend, doch hat die Theorie dafür bisher keine überzeugende Formulierung vorgelegt.

Allgemein gilt, daß Änderungen solcher Kognitionen nicht erfolgen, die selbst wiederum Dissonanz entstehen lassen würden. Entsprechend wird vermutet, daß selbstbezogene Kognitionen sich nicht ändern werden. So zeigt die Untersuchung von ARONSON & CARLSMITH (1962), daß zum Ausgleich von überhöhter Fremdbewertung und Selbstkonzept die Leistung verringert wird,

um die Fremdbewertung dem Selbstkonzept anzugleichen. Mehrfache Replikationen der Untersuchung haben das Ergebnis aber nicht bestätigen können (IRLE 1975). IRLE & KROLAGE (1973) finden entsprechend, daß von der Selbstbewertung abweichende Fremdbewertungen (als Mitteilungen von Intelligenztestergebnissen), die zu Dissonanzen führen müssen, durchaus über Änderungen des Selbstkonzepts ausgeglichen werden. Die Änderung nimmt zu mit der Stärke der Abweichung. Zusätzlich zu dem Ausgleich über die Änderung des Selbstkonzepts sind noch weitere Lösungen der Dissonanz zwischen den Kognitionen durch fehlerhafte Erinnerungen der Resultate, durch Änderung der Beurteilung von Vergleichspersonen, durch Änderung der Einschätzung der Validität der verwendeten Testverfahren usw. festzustellen. Die Dissonanztheorie ist nicht in der Lage, die spezifische Form der Änderung vorherzusagen; sie zeigt allein auf, daß psychische Kräfte auf Herstellung von Konsonanz drängen. Es wird von konkreten situativen Umständen abhängen, ob und gegebenenfalls wie die entstandene Dissonanz über eine Veränderung des Selbstkonzepts verringert werden kann.

Die Arbeit von IRLE & KROLAGE (1973) weist aber als eine Reaktion auf abweichende Fremdbewertungen auch eine solche aus, die konsistenztheoretisch nicht gut zu deuten ist: Änderungen erfolgen in positiver Richtung stärker als in negativer Richtung. Das kann mit Befunden in Zusammenhang gebracht werden, die aus solchen Untersuchungen bekannt sind, in denen die Einstellung zum Bewerter in Abhängigkeit von der Beziehung Selbstkonzept-Fremdbewertung sich ändert. Bei DEUTSCH & SOLOMON (1959) ist neben einem konsistenztheoretisch zu deutenden Effekt auch ein "Positivitäts-Effekt" zu beobachten: Positive Fremdbewerter werden mehr geschätzt. Eine Wiederholung der Untersuchung von DEUTSCH & SOLOMON durch SKOLNICK (1971) mit Studenten

als Versuchspersonen, während bei DEUTSCH & SOLOMON Arbeiterinnen als Versuchspersonen dienten, findet nur Reaktionen in der Richtung des Positivitäts-Effekts. Darauf baut eine Theorie auf, die im Gegensatz zur Konsistenztheorie steht: die Theorie des Selbstwertgefühls (JONES 1973). Weniger noch als bei der Dissonanztheorie ist der erhobene Theorieanspruch voll gerechtfertigt, da kaum mehr als ein zentraler Begriff vorliegt. Die Theorie des Selbstwertgefühls bezieht sich vorrangig auf die beschriebene Einstellung der bewerteten Person zu dem Bewerter. Sie erwartet, daß diese Einstellung allein von der Positivität der Bewertung abhängig ist und nicht, wie Konsistenztheorien vorhersagen, von der Konsistenz der Fremdbewertung mit dem entsprechenden Selbstkonzept. Das hat noch keine direkte Beziehung zur Wirkung der Fremdbewertung auf das Selbstkonzept. Man kann jedoch vermuten, daß die Einstellung zu dem Bewerter die Bereitschaft zur Änderung des eigenen Selbstkonzepts beeinflußt. JONES diskutiert dieses Problem nicht, sondern setzt an bestimmten Stellen seiner Arbeit beide Prozesse einfach einander gleich, was eine weder theoretisch noch empirisch zureichende Lösung darstellt.

Beide Theorien ermöglichen in dieser abstrakten Gegenüberstellung noch keinen Hinweis auf die wahrscheinlichen Änderungen in der vorliegenden Situation. Es wird von den konkreten Situationsbedingungen abhängen, in welcher Richtung die Reaktion erfolgt; außerdem sind, wie die Ergebnisse von IRLE & KROLAGE (1973) zeigen, Mischungen beider Reaktionen möglich. GERGEN (1971) versucht, allerdings im Hinblick auf die Einstellung zum Bewerter als abhängige Variable, die situativen Momente zu bestimmen, die über die Reaktion im Sinne einer der beiden Theorien entscheiden. Für die Situation der Fremdbewertung über Noten in Prüfungen ergibt sich daraus, daß sowohl konsistenztheoretisch

begründete Reaktionen erfolgen sollten (der konditionale Auslesecharakter der Situation spricht dafür) als auch selbstwerttheoretisch begründete (die persönliche Beziehung und die persönliche Bedeutung der Prozesse sprechen dafür).

Insgesamt gesehen kann diese Analyse jedoch nicht befriedigen. Sie ist u.a. dadurch belastet, daß sie primär auf die Reaktionen gegenüber dem Bewerter hin angelegt ist und nicht auf Änderungen eigener Bewertung. Sie deutet aber an, daß eine genauere Analyse der Situation erst konkrete Aussagen ermöglicht. Nur am Rande sei vermerkt, daß die fehlende Berücksichtigung von situativen Bedingungen und darüber hinaus von sozialen Rahmenbedingungen diese "sozial"psychologischen Theorien insgesamt zur erklärenden Bewältigung komplexer sozialer Realität untauglich macht. Wenn die soziale Lebensumwelt der Versuchspersonen nur als Störbedingung erwähnt wird (SKOLNICK 1971) anstatt in die Theorie integriert zu werden, muß ein derartiges Theoretisieren die soziale Realität systematisch verfehlen.

Die Analyse <u>situativer Bedingungen</u> stützt sich auf empirische Untersuchungen, die nicht nur dem Rahmen der beiden erwähnten Theorien entstammen. Sie lassen sich nach einzelnen Kennzeichen ordnen, die Bestimmungsstücke der Situation darstellen: Glaubwürdigkeit des Bewerters (und der Bewertung), Diskrepanz der Bewertung, Konsistenz der Bewertung, Positivität der Bewertung, Öffentlichkeit des Selbstkonzepts, Möglichkeit von Alternativreaktionen. Teilweise sind für einzelne der situativen Bestimmungsstücke noch Interaktionen nachweisbar, die aber nicht systematisch erörtert werden.

Die Glaubwürdigkeit der Bewertung, als Glaubwürdigkeit des Bewerters und des Prozesses der Bewertung zu sehen, hat einen Einfluß auf die Änderung des

Selbstkonzepts: Bei höherer Glaubwürdigkeit nimmt die Änderung zu (GERGEN 1971). Allein eine größere Bekanntheit des Bewerters reicht zur Erhöhung des Einflusses jedoch nicht aus (HARVEY et al. 1957).

Das Ausmaß der Änderung hängt von der Diskrepanz der Fremdbewertung zum bestehenden Selbstkonzept ab: Mit größerer Diskrepanz ist das Ausmaß der Änderung größer (IRLE & KROLAGE 1973). Jedoch verringert sich, besonders bei negativ abweichender Fremdbewertung, die Konformität ab einem bestimmten Ausmaß der Diskrepanz wieder (JOHNSON 1966), was mit einer Abnahme der Glaubwürdigkeit bei zu großer Diskrepanz zusammenhängen kann.

Werden die Fremdbewertungen von einer Gruppe von Bewertern abgegeben, ist die Änderung des Selbstkonzepts größer, wenn die Bewerter übereinstimmen (SHERWOOD 1967). Eine von der Gruppe abweichende Bewertung wird auch dann nicht berücksichtigt, wenn sie für die bewertete Person positiv ist (HOWARD & BERKOWITZ 1958).

Die bisher berichteten Ergebnisse sind durchweg nicht unwidersprochen: Unter bestimmten Bedingungen treten auch andere als die dargestellten Zusammenhänge auf. Als weitgehend universell dagegen erweist sich der Effekt der Richtung der Bewertung: ob die Fremdbewertungen einen Erfolg oder einen Mißerfolg aussprechen. WYLIE (1961) faßt die Untersuchungsergebnisse dahingehend zusammen, daß das Selbstkonzept nach Erfolg eher steigt als nach Mißerfolg sinkt. Das Ergebnis ist nicht ohne Einschränkung durch die Theorie des Selbstwertgefühls zu vereinnahmen; es macht vielmehr einen Widerspruch in dieser Theorie, wenn sie auf Änderungen des Selbstkonzepts angewendet wird, deutlich. Bei den Experimenten, auf die sich WYLIE beruft, wie z.B. auch bei späteren Untersuchungen von IRLE & KROLAGE (1973), SOLWAY & FEHR (1969) handelt

es sich bei den Fremdbewertungen um einmalige Ereignisse. Würde die stärkere Berücksichtigung von Erfolg durchgehend der Fall sein, müßten Selbstbewertungen zunehmend über die wirklichen Fähigkeiten hinauswachsen; als allgemeiner Trend ist der Positivitäts-Effekt deshalb nicht anzunehmen. Da Prüfungen aber in der vorliegenden Form, den Experimenten ähnlich, isolierte Ereignisse sind, ist bei ihnen ein Positivitäts-Effekt zu erwarten.

Eine Korrektur erfährt die Feststellung, daß nach Mißerfolg weniger entsprechende Änderungen des Selbstkonzepts zu beobachten sind, durch einen sorgfältigen, aber nicht noch einmal aufgenommenen Ansatz von DILLER (1954): Auf Erfolg hin erfolgen offene, auf Mißerfolg hin nur verdeckte, projektive Änderungen des Selbstkonzepts. Damit ist die soziale Dynamik der Befragungssituation angesprochen. Schon anläßlich möglicher Erwünschtheitstendenzen wurde darauf verwiesen, daß dem Selbstkonzept in diesem Rahmen durchaus die soziale Dynamik der Situation der Äußerung des Selbstkonzepts als valider Bestandteil zuzurechnen ist. Insofern hat die Korrektur durch DILLER keine besondere Bedeutung.

Die Bedeutung der sozialen Dynamik der Äußerung des Selbstkonzepts wird auch in Untersuchungen zur Öffentlichkeit der Situation aufgezeigt. Schon die geringfügige Andeutung der Öffentlichkeit des Selbstkonzepts kann einen Einfluß haben (IRLE 1975). Erwartete Öffentlichkeit als Aussicht auf eine Überprüfung des Selbstkonzepts verhindert eine Erhöhung nach positiv abweichender Fremdbewertung (EAGLY & ACKSEN 1971); erwartete Veröffentlichung der mitgeteilten Bewertung führt zu einer Verstärkung der Reaktion auf positive Abweichungen (GERARD 1961).

Als letzte Bedingung, die nicht mehr rein der Situation zuzurechnen ist, sind die Alternativreak-

tionen zu nennen, die bei negativer Abweichung der Fremdbewertung möglich sind. Sie nehmen mit der negativen Abweichung der Fremdbewertung zu und ändern sich dabei in ihrer Charakteristik (JOHNSON 1966). Für die vorliegende Situation ist ausschlaggebend, ob diese Reaktionen Alternativen oder nur Parallelerscheinungen darstellen. Bei HARVEY et al. (1957) ändert sich das Selbstkonzept auf eine negative Fremdbewertung hin trotz gleichzeitiger Abwertung, Verzerrung usw. dieser Bewertung; auch bei IRLE & KROLAGE (1973) finden die Alternativreaktionen gleichzeitig mit Änderungen des Selbstkonzepts statt. Doch IRLE (1975) weist in einer weiterführenden Untersuchung nach, daß wirkliche Alternativen möglich sind. Das bestätigt JOHNSON (1966), der mit der Zunahme der negativen Diskrepanz weniger Änderung des Selbstkonzepts und bedeutsamere Alternativreaktionen feststellt. Für die zumindest teilweise Ausschließlichkeit von Selbstkonzeptänderung und Alternativreaktionen spricht die Abnahme der Glaubwürdigkeit als eine Form von Alternativreaktion, die gleichzeitig eine Bedeutung für die Änderung hat (s. oben).

Neben den dargestellten einzelnen Effekten sind, wie angedeutet, vielfältige Interaktionen möglich. Beispielsweise sei auf die Interaktion von Glaubwürdigkeit und Diskrepanz hingewiesen: Bei hoher Glaubwürdigkeit nimmt die Änderung mit der Größe der Diskrepanz zu; bei geringer nimmt sie mit steigender Diskrepanz ab (BERGIN 1962).

Die Übertragung dieser Ergebnisse auf die vorliegende Situation kann nur mehr oder weniger summarisch erfolgen, da einzelne Situationsmomente nicht genügend genau identifizierbar sind (s. oben). Sie hat zudem prinzipiell zu bedenken, daß ein weitgehender Unterschied zwischen der auf bestimmte Handlungs- und Erlebnismöglichkeiten der Personen eingeschränkten experi-

mentellen Situation und der Situation der Prüflinge besteht, in der die Personen auf komplexe Art und Weise die Fremdbewertung verarbeiten können. Es kann vermutet werden, daß die Bedeutung einzelner verfälschender situativer Faktoren und ihre Interaktion eher größer sein werden, als in Experimenten nachzuweisen ist.

Die Übertragbarkeit der in den experimentell angelegten Untersuchungen gewonnenen Ergebnisse scheint ferner von einer grundsätzlichen Bestimmung der vorliegenden Situation abhängig. Nach TEDESCHI (1972) kann man bei sozialem Einfluß zwischen normativem und informativem Einfluß unterscheiden. Im Fall von informativem Einfluß ist die fremde Person im wesentlichen nur Mitteiler eines Ergebnisses, im Fall von normativem wird sie als Quelle von Belohnungen und Bestrafungen wahrgenommen (GERARD 1961). JONES (1973) versucht, diese Unterscheidung mit dem Begriff des Personalismus in seiner Theorie zu berücksichtigen. Nun stellen sowohl viele der experimentellen Situationen als auch die Prüfungssituation eher Mischungen dar, in denen die Bewertung durch eine andere Person nicht ohne Grundlage erfolgt, aber auch ein Bewertungsinstrument nicht für sich spricht, sondern erst über die kompetente Auslegung durch eine andere Person. Dadurch sind Versuchssituationen solcher Art selten frei von Autoritätsbeziehungen. Die oben berichteten Situationsmomente wie Glaubwürdigkeit, Diskrepanz, Konsistenz, Positivität und Öffentlichkeit sind ebenfalls nicht klar den beiden Einflüssen zuzuordnen. Ein prinzipieller Einwand gegen die Übertragung ergibt sich aus dieser Unterscheidung deshalb nicht.

Nicht nur in der Situation, sondern auch in der Person sind Bedingungen aufzuzeigen, die den Prozeß der Beeinflussung unterschiedlich gestalten. Es wurde schon darauf hingewiesen, daß das bestehende Selbst-

konzept eine Bedingung für die Reaktion auf Fremdbewertungen ist. Positiv ausgeprägte Selbstkonzepte sind in stärkerem Maße gegen eine Änderung resistent als weniger positive (SKOLNICK 1971). Über das jeweilige Selbstkonzept hinausgehend hängt die Beeinflußbarkeit vom übergreifenden Selbstwertgefühl ab: Bei einem insgesamt positiven Selbstwertgefühl ist weniger Einfluß auf das Selbstkonzept einer Person möglich (GERGEN 1971). Ein positives Selbstwertgefühl schützt insbesondere vor der Verarbeitung von negativer Bewertung (STOTLAND et al. 1957).

Als weitere wesentliche persönlichkeitsspezifische Bedingungen des Einflusses stellen sich u.a. die folgenden dar: die Ängstlichkeit der Person (DORIS & SARASON 1955; DORIS 1959); das individuelle Anspruchsniveau (HOWARD & BERKOWITZ 1958); die erlebte Selbstverantwortlichkeit (internal/external locus of control; WOLFE 1972).

Die letzte Untersuchung hat Vorhersagen eigener Studienleistungen zum Gegenstand, die wenn, dann nur einen Aspekt des Selbstkonzepts ansprechen. Sie ist aber deshalb von Interesse, weil sie eine Interaktion von Persönlichkeitsbedingungen und sozialen Bedingungen belegt. Nur bei Studienanfängern hat die erlebte Selbstverantwortlichkeit einen Einfluß auf die Verarbeitung der Fremdbewertung. Eine Interaktion von übergreifendem Selbstwertgefühl und Richtung der Fremdbewertung ist Ergebnis einer Untersuchung von SCHNEIDER (1969).

Damit ist die Auswertung von Erkenntnissen sozialpsychologischer Grundlagen situativen Einflusses auf das Selbstkonzept abgeschlossen. Die Darstellung läßt erkennen, daß zwischen experimentellen Situationen und den Verhältnissen in und nach Prüfungen große Unterschiede bestehen. Deshalb sollen Erkenntnisse aus Situationen, die eher mit der Prüfung vergleichbar sind, zusätzlich zur Abklärung möglicher Wirkungen der

Noten herangezogen werden. Eine solche Situation ist die Mitteilung diagnostischer Information.

Einfluß diagnostischer Information

Es ist bezeichnend für den fremdbestimmenden Charakter der Psychologischen Diagnostik im klinischen und pädagogischen Bereich, daß wenig Untersuchungen über die Auswirkungen der Mitteilung diagnostischer Informationen durchgeführt werden. Entsprechend beschränken sich praktische Hinweise für die adäquate Gestaltung von Mitteilungssituationen (APPELBAUM 1970) auf Banalitäten. Im Rahmen einer zunehmend testkritischen Wende deutet sich in den USA eine Änderung an. Die zusammenfassenden Berichte von GOSLIN (1972) und KIRKLAND (1972) bestätigen im wesentlichen jene situativen und personalen Bedingungen und deren Interaktionen, die bereits im vorigen Abschnitt als wirksam herausgestellt wurden. Nach GOSLIN sind die Glaubwürdigkeit von Test und Bewerter, die Konsistenz der diagnostischen Information und die Greifbarkeit von Rationalisierungen für eine angemessene Wirkung ausschlaggebend. Die Verarbeitung negativer Information wird als problematisch herausgestellt.

Einige Untersuchungen mögen das belegen: Sie beschränken sich auf solche Situationen, in denen eine reine Mitteilung der Ergebnisse erfolgt (wie das auch in der Prüfung der Fall ist). Ergebnisse über Auswirkungen von Beratungen, in die die Ergebnismitteilung eingebettet wird, werden in Kapitel IV erörtert. Die Mitteilung von Testergebnissen, die vom Selbstkonzept der Probanden abweichen, hat Änderungen des Selbstkonzepts im Sinne dieser Testergebnisse zur Folge (IRLE & KROLAGE 1973; WESTBROOK 1967). Eine Änderung erfolgt eher in positiver Richtung nach positiver Abweichung als in negativer Richtung nach negativer Abweichung (BRIM 1967; COMPTON 1941).

Über die allgemein sozialpsychologischen Erkenntnisse des vorigen Abschnitts hinausgehend machen Untersuchungen zur Rückmeldung von Testergebnissen deutlich, daß es sich bei der Zuschreibung der Wirkung zur Mitteilung der Ergebnisse um eine "Scheinrealisation" (HOLZKAMP 1968) handeln kann: Schon die reine Absolvierung eines Tests bedeutet in bestimmtem Ausmaß eine Rückmeldung und kann unabhängig von der Mitteilung in Form einer darauf aufbauenden Bewertung einen Einfluß haben. So ist bei ARSENIAN (1942), FROEHLICH (1954; 1957), ROSS (1933) eine, wenn auch geringfügige, Zunahme der Realistik des Selbstkonzepts allein durch die Testdurchführung festzustellen. Auf den gleichen Effekt macht AMMONS (1956) für einzelne Lernakte aufmerksam.

Der Bericht von GOSLIN (1972) betont einen Aspekt, der schon kurz angesprochen wurde, aber weder in der bisherigen Forschung beachtet wird, noch in der empirischen Untersuchung im Rahmen dieser Arbeit: Bei der Rückmeldung von Testergebnissen kann nicht nur die Abweichung des Ergebnisses vom Selbstkonzept der Person einen Einfluß haben, sondern gleichfalls auch die Nichtabweichung als Bestätigung des Selbstkonzepts.

Einfluß von Noten

Der Rückgriff auf Ergebnisse aus anderen Bereichen in den vorangehenden beiden Abschnitten war notwendig, weil nur wenige Untersuchungen über die Wirkung von Noten vorliegen. Da die Situationen, in denen die Noten erteilt werden und wirksam werden können, keineswegs mit der hier untersuchten Verwendung der Noten übereinstimmen, darf die aus den folgenden Untersuchungen gewonnene Erkenntnis nicht überbewertet werden.

HETZER (1933) berichtet in fallstudienartigen Erfahrungen, daß die Entwicklung der Selbstbewertung

bereits bei 13 bis 18jährigen Schülern sich zunehmend von den Zeugnissen ablöse. Außerdem glaubt sie, eine geschlechtsspezifische Wirkung feststellen zu können; eine Bedingung, die bisher zwar im Hinblick auf die Höhe des Selbstkonzepts, aber nicht im Hinblick auf Änderungsprozesse angesprochen wurde.

Im Rahmen von ausführlichen Befragungen zum Problem der Verwendung von Testverfahren stellen BRIM et al. (1965) und BRIM (1966) fest, daß die in der schulischen Ausbildung erteilten Noten von Schülern der High-School als wesentlichste Quelle des Selbstkonzepts eigener Intelligenz angegeben werden. Nur Schüler in Eliteschulen orientieren sich mehr an den Urteilen ihrer Bezugspersonen. Von amerikanischen Erwachsenen insgesamt werden die Noten als zweitwichtigste Quelle der Einschätzung ihrer eigenen Intelligenz betrachtet. Dabei sind die Noten bedeutsamer bei jüngeren Personen, bei weiblichen Personen, bei Personen weißer Hautfarbe sowie bei Personen mit mehr Erziehung und einer höheren Position.

Es konnten bei der nachträglichen Literaturauswertung nur zwei Untersuchungen ermittelt werden, die einen in etwa dieser Untersuchung vergleichbaren Ansatz verfolgen (BOSHIER 1972; FRETZ & ENGLE 1973). FRETZ & ENGLE lassen in einem Psychologiekurs eine Testprüfung absolvieren, nach der nur ein Teil der Prüflinge eine Rückmeldung über Buchstabennoten und Abweichungswerte erhält. Die Ergebnisse der Untersuchung werden unvollständig berichtet. Es wird aber deutlich, daß die Rückmeldung nur insignifikant mehr Änderungen des Selbstkonzepts bewirkt. Dabei ist der Unterschied zwischen Studienanfängern und fortgeschrittenen Studenten bedeutsam: Anfänger ändern ihr Selbstkonzept mehr. Wird die Rückmeldung in Beziehung zur bisherigen Leistung oder zur Leistungserwartung gesetzt, zeigt die Abweichung keinen Zusammenhang mit

dem kursspezifischen Selbstkonzept, aber eine komplexe Interaktion in bezug auf das globale Selbstkonzept. Letzteres ändert sich zumindest bei Anfängern signifikant stärker, wenn die Diskrepanz zur vergangenen Leistung groß ist. Durch den lückenhaften Bericht der Ergebnisse wird eine Einschätzung erschwert. Die Verfasser betonen, daß im wesentlichen komplexe Interaktionen das Bild der Wirkung der Rückmeldung durch Noten kennzeichnen.

BOSHIER manipuliert die Rückmeldung in einer wichtigen Zwischenprüfung, so daß ein Teil der Prüflinge einen die Zulassung zur Abschlußprüfung ausschließenden Mißerfolg erfährt. Nach diesem Mißerfolg zeigen die Prüflinge keine bedeutsamen Änderungen ihrer Werte in einer Messung des globalen Selbstkonzepts gegenüber einer sechs Wochen vor der Zwischenprüfung erfolgten Messung. Die Diskrepanzmessung des Selbstkonzepts ergibt nach Mißerfolg sogar signifikant positivere Werte. Der Autor deutet die Ergebnisse mit der Abwehr von Mißerfolg zur Erhaltung des Selbstwertgefühls; die Ergebnisse bei der Diskrepanzmessung lassen ihn an der Bedeutung solcher Diskrepanzwerte zweifeln (s.oben Kapitel III.I.C.).

Die beiden Untersuchungen gehen mit der Verwendung von globalen Messungen des Selbstkonzepts von einem mehr persönlichkeitspsychologischen Ansatz aus; sie sind deshalb, obwohl die Situation eher den hier gegebenen Prüfungen vergleichbar ist (bei BOSHIER wegen der alleinigen Verwendung von Mißerfolg allerdings weniger), doch nicht aussagekräftiger als Untersuchungen der vorangehenden Abschnitte, die spezifischere, fähigkeitszentrierte Messungen des Selbstkonzepts verwenden.

Situationsbezogene Integration der Ergebnisse

Die situationsbezogene Integration der Ergebnis-

se soll abschätzen, wie die Wirkung der Noten in Prüfungen im Vergleich mit den gesetzten Normwerten ausfallen wird. Diese Voraussicht bezieht sich, das sei noch einmal betont, nicht nur auf die in Teil B dargestellte empirische Untersuchung, sondern hat auch unabhängig davon, ob deren Ergebnisse in ihrem Sinne ausfallen, Bedeutung, weil sie auf zwar weniger situationsangepaßte, aber aspektreichere (und gleichfalls empirische) Erfahrung aufbauen kann. Für die Anwendung dieser Erfahrung ist über die eingestreuten Bemerkungen hinausgehend abzusichern, daß sie auf die vorliegende Situation passen.

Die Integration kann in zwei Schritten erfolgen. Ein erster Schritt ist durch den Text schon weitgehend getan: Es ist sichergestellt, daß Noten in Prüfungen Fremdbewertungen für Fähigkeiten darstellen und in der Mischung von informativen und bewertenden Komponenten in Experimenten verwendeten Fremdbewertungen gleichen. Insofern ist eine Wirkung im Sinne der Normwerte zu erwarten, die aber durch viele situative und personale Bedingungen und deren Interaktionen nicht allzu deutlich ausfallen dürfte, sondern mannigfache Verzerrungen als Abweichungen von einer realistischen Abbildung aufweist. Zur Bestätigung dieser Erwartung, die der Verwendung von Noten zur Rückmeldung allgemein gilt, können die Aussagen von Texten der Pädagogischen Psychologie herangezogen werden, die in ihrer Gesamtheit gleichfalls eine eingeschränkte Wirkung vermuten, wenn auch einzelne Texte (nicht zuletzt infolge nicht explizierter Erwartungen) zu unterschiedlichen Urteilen kommen. So wird behauptet, daß sich die Noten als Information für die Lernenden durchaus eignen (DAVIS 1964), es wird die Einfachheit und Verständlichkeit der Noten betont (DOHSE 1967; HASEMANN 1970b; RINSLAND 1938). Demgegenüber stehen andere Texte, die eine Eignung der Noten für die Rück-

meldefunktion bezweifeln: Das wird aus der zu geringen Differenzierung der Noten begründet (AHMANN & GLOCK 1967; BROWN 1971). Andere Texte verweisen auf die Unpersönlichkeit, Allgemeinheit und fehlende Interaktionsmöglichkeit als Nachteile der einfachen Mitteilung (THORNDIKE 1972) und empfehlen eine Ergänzung durch zusätzliche Hinweise (LIEN 1967).

Ein zweiter Schritt aber muß versuchen, über diese immer noch allgemeinen Bemerkungen hinausgehend die konkrete Besonderheit der vorliegenden Prüfungssituation einzubeziehen. Das scheitert, wie anfänglich angedeutet, an fehlender Kenntnis der Bedingungen, doch kann zumindest an einigen Bedingungen angesetzt werden; eine systematische Abarbeitung aller dargestellten Bedingungen ist nicht beabsichtigt.

Die Glaubwürdigkeit von Prüfungen und Noten, die eine situative Bedingung ihrer Abbildung im Selbstkonzept darstellt, ist nicht besonders hoch. Zum einen dürften die konkreten Eigenschaften der Noten und Prüfungen, die in Kapitel III.I.B. berichtet werden, den Prüflingen nicht verborgen bleiben. Zum anderen beweist die Untersuchung der Einstellung zur Prüfung (TIMMERBRINK 1971), daß die Prüflinge dem Informationsgehalt von Noten und Prüfungen sehr skeptisch gegenüberstehen.

Die naheliegende Abwertung der Zuverlässigkeit und Validität der Noten ist zugleich eine der möglichen Alternativreaktionen, die eine Verarbeitung diskrepanter Information umgehen läßt. Die Situation nach der Prüfung stellt darüber hinaus in ihrer Komplexität sehr viel mehr alternative Reaktionen bereit, als das in reduzierten Experimentalsituationen der Fall ist. Außerdem ist die Verarbeitung selbst kein rein individueller Prozeß, sondern schließt vielfältige soziale Interaktionen, individuell unterschiedlich, mit ein.

Spricht dies vermehrt gegen eine normgerechte

Wirkung der Noten, ist auch eine eher fördernde Bedingung auszumachen. Es fehlt ein Situationsmoment weitgehend, das sich einer normgerechten Wirkung widersetzt: die Rückmeldung von Mißerfolg. Die Noten in der vorliegenden Prüfung stellen ihrer empirischen Verteilung nach vermutlich selten einen klaren Mißerfolg dar (s. auch Kapitel III.III.). Die Normangemessenheit der daraus erwarteten Wirkung aber ist nicht ohne Problematik. Zwar mag das Fehlen von Mißerfolg zu einer realistischen Abbildung der relativen Leistungen beitragen dadurch, daß das Ausmaß positiver Diskrepanz realistisch abgebildet wird. Auf der anderen Seite kann es aber die schon bestehende Überschätzung noch weiter verstärken.

Die bisherigen Überlegungen gehen allesamt davon aus, daß die vorliegende Situation in ihren Grundzügen mit den Situationen vergleichbar ist, in denen Fremdbewertungen im Rahmen von Experimenten mitgeteilt werden. Es ist jedoch eine bedeutsame Abweichung festzustellen: Die Situationen in den ausgewerteten Untersuchungen sind durchweg so beschaffen, daß die Rückmeldung den wesentlichen Inhalt der Situation darstellt. Bei der Rückmeldung von Noten in diesen Prüfungen aber ist die Rückmeldung nur ein Randaspekt, da die Noten fremdbestimmend unabhängig von der Rückmeldung wirksam werden können. Das kann sowohl zusätzliche Alternativreaktionen aktivieren; es kann auch einen Zwang zur realistischen Reaktion darstellen, wie die Untersuchung von GERARD (1961) es nahelegt (dort bleibt dieser Effekt bei negativen Diskrepanzen aus, die in dieser Prüfung jedoch auch selten sein dürften; s. oben).

Zusammenfassung (III.II.A.1.)

Selbstkonzepte sind als Sozialisationsprodukte prinzipiell sozial beeinflußbar. Die Beeinflußbarkeit

von Selbstkonzepten, die auf eine bestimmte einmalige Situation zurückgeht, wird im Rahmen der Dissonanztheorie und der Theorie des Selbstwertgefühls gesehen. Beide kommen zu keinen klaren Ergebnissen: Auf Grund der Dissonanztheorie ist eine Änderung im Sinne der Normwerte möglich, die Theorie des Selbstwertgefühls schließt eine solche eher aus. Die von den Theorien relativ losgelöste Analyse der situativen Bedingungen des Einflusses auf das Selbstkonzept zeigt, daß verschiedene Bedingungen bedeutsam sind. Die Wirkung hängt ab von der Glaubwürdigkeit, der Diskrepanz, der Konsistenz und der Öffentlichkeit der Bewertung sowie vom Einsetzen von Alternativreaktionen bei negativen Bewertungen. Nach Erfolg ergeben sich eher Wirkungen als nach Mißerfolg. Zusätzlich gehen die Bedingungen Interaktionen miteinander ein.

Neben den situativen sind an möglichen Wirkungen auch personale Bedingungen wie das umfassende Selbstwertgefühl, das Anspruchsniveau, die Ängstlichkeit usw. beteiligt. Untersuchungen aus dem Bereich der Mitteilung diagnostischer Informationen bestätigen die Bedingungsabhängigkeit der Wirkung im wesentlichen. Einflüsse im Sinne der Normwerte werden durch eine Reihe von verfälschenden situativen und personalen Bedingungen überlagert. Die insgesamt seltenen Untersuchungen an Noten machen deutlich, daß ein Einfluß der Noten in bedeutsamem Ausmaß nicht zu erwarten ist.

Die abschließende Integration der Erkenntnisse geht davon aus, daß die Ergebnisse von Untersuchungen aus anderen Bereichen auf die vorliegende Situation zu übertragen sind. Es ist deshalb mit einer Wirkung, die in der durch die Normwerte vorgezeichneten Richtung liegt, zu rechnen, wenngleich die Wirkung durch die Bedeutung von Situations- und Persönlichkeitsbedingungen und deren Interaktionen (auch miteinander) beeinträchtigt wird. Bestimmte Eigenschaften der konkreten

Situation wie deren geringe Glaubwürdigkeit und die Tatsache, daß die Rückmeldung nicht als der eigentliche Zweck der Situation erlebt werden kann, sprechen zusätzlich gegen eine normgerechte Wirkung, während das Ausbleiben von deutlichem Mißerfolg eine fördernde Bedingung darstellt.

III.II.A.2.: Wirkungen auf die Motivation

Die folgenden Ausführungen haben implizit eher den Aspekt der Handlungsplanung im Auge, da die Beliebtheit nur als Kontrollfrage fungieren soll. Dabei fällt es schwer, wie die Erörterung des theoretischen Hintergrundes zeigte, den Prozeß der Handlungsplanung im Kontext der herkömmlichen Motivationspsychologie zu sehen. Dementsprechend bereitet auch die Sammlung von Erkenntnissen über die Wirkung von Fremdbewertungen auf die Änderungen der Handlungsplanung Schwierigkeiten. Ist immer schon die Übertragung von Erkenntnissen durch Unterschiede zwischen den Situationen der Erkenntnisgewinnung und der vorliegenden Situation belastet, tritt hier noch die nur partielle Identität des Gegenstands bestehender Theorien bzw. einzelner empirischer Untersuchungen mit dem Gegenstand der Frage hinzu.

Einfluß im Modell der Leistungsmotivation

Um die Theorie der Leistungsmotivation als Erkenntnisgrundlage heranziehen zu können, ist eine Vorannahme notwendig, die besagt, daß die Aussagen der Leistungsmotivationsforschung noch gültig sind, wenn die Motivation um bestimmte gegenständlich differenzierte Bezüge (Einzelfächer) angereichert wird. Eine zweite Vorannahme geht davon aus, daß die an kurzzeitig erstreckten Person/Umwelt-Bezügen gewonnenen Erkenntnisse (WEINER 1976) zumindest dem Prinzip nach

auf zeitlich langerstreckte Bezüge und deren kognitive Antizipation, wie sie Handlungsplanungen darstellen, übertragbar sind.

Im klassischen Modell der Leistungsmotivation (ATKINSON 1964) kann ein situativer Einfluß nur über die subjektive Erfolgswahrscheinlichkeit vermittelt werden. In vielen Experimenten hat sich herausgestellt, daß diese Erfolgswahrscheinlichkeiten durch sozialen Einfluß, u.a. auch durch die Suggestion des Versuchsleiters, beeinflußt werden können. Jedoch ist die Annahme, daß die Motivation bei mittleren Erfolgswahrscheinlichkeiten am größten sei, in dieser Einfachheit kaum noch zu halten. Die Beziehung von Erfolgswahrscheinlichkeit und Motivation ist selbst abhängig von der Struktur individueller, situativ überdauernder motivischer Tendenzen der Person (MEYER 1973). Für komplexe Aufgaben ist zudem zu berücksichtigen, daß die Optima der Erfolgswahrscheinlichkeit infolge von Verknüpfungen mit zeitlich weiterstreckten Leistungszielen bei sehr viel niedrigeren Erfolgswahrscheinlichkeiten als den ursprünglich angenommenen mittleren Wahrscheinlichkeiten liegen können (RAYNOR 1968). Da über die Erfolgswahrscheinlichkeiten in der vorliegenden Situation nichts bekannt ist, sind keine Aussagen über die Richtung möglicher Änderungen der Motivation möglich. Es bleibt aber festzuhalten, daß eine einfache Abbildung vermutlich durch personal stabile Tendenzen "gestört" wird.

Direkte empirische Hinweise auf eine mögliche Wirkung der Rückmeldung über die Erfolgswahrscheinlichkeit geben die Untersuchung von ESKOLA (1961), die einen Zusammenhang zwischen der Qualität der Abschlußnote des Schulbesuchs und der Wahl der Schwierigkeit von Kursen an der Hochschule berichtet, und die Untersuchung von FLOOK & SAGGAR (1968), die eine leistungssteigernde Wirkung der Rückmeldung von Testergebnissen

an Studenten durch die dadurch ermöglichte realistischere Erfolgserwartung erklären. Allerdings ist die letztere Untersuchung nicht ohne Widerspruch geblieben. FLOOK & ROBINSON (1972) finden in einer Wiederholung, die außerdem die Schätzung der subjektiven Erfolgswahrscheinlichkeiten einbezieht, keinen Einfluß auf diese Wahrscheinlichkeiten und einen nur unbedeutenden Einfluß auf die Leistungserhöhung allein im mittleren Leistungsbereich.

Unter der Annahme, daß die Zuschreibung von Erfolg und Mißerfolg eine kognitive Vorstufe der erlebten Erfolgswahrscheinlichkeit für weitere Aktivitäten darstellt, kann die Untersuchung von SIMON & FEATHER (1973) als Beweis einer zusätzlichen "Störvariablen" herangezogen werden: Nach Erfolg ist die Attribution anders gerichtet als nach Mißerfolg. Das könnte dann auch bedeuten, daß die subjektiven Erfolgswahrscheinlichkeiten nicht den wirklichen Rückmeldungen angepaßt werden müssen.

Insgesamt gesehen ergibt sich aus der Analyse möglicher Änderungen der subjektiven Erfolgswahrscheinlichkeiten nur wenig positive Evidenz für eine normgerechte Wirkung der Fremdbewertung auf die Motivation. Die Aussage aber muß insofern relativiert werden, als die Leistungsmotivationsforschung selbst solchen Problemen situativen Einflusses wenig Aufmerksamkeit widmet.

Die letzte Einschränkung gilt noch mehr für ein weiteres Konzept im Rahmen der Leistungsmotivationsforschung, das als über die Handlung hinausweisende motivationale Tendenz der auf die Prüfung folgenden Handlungsplanung entsprechen könnte: der Trägheitsmotivation (WEINER 1976). Das Konzept spricht für die ursprüngliche Annahme eines kurvilinearen Zusammenhangs zwischen Rückmeldung und Handlungsplanung, indem es davon ausgeht, daß motivationale Energien nach Miß-

erfolg erhalten bleiben. In dieser Beschränkung auf Mißerfolg und empirische Hinweise, die bisher über Prozesse kurzzeitiger Dauer nicht hinausgehen, läßt das Konzept wenig Bedeutung für die vorliegende Situation erkennen. (Allerdings weisen die Ergebnisse von FISCH (1970), der bei Prüflingen auch noch drei Wochen nach der Prüfung besondere leistungsthematische Aktivierungen findet, auf eine universellere Bedeutung der Trägheitsmotivation hin.)

In nur noch losem begrifflichen Zusammenhang mit der Leistungsmotivationsforschung stehen Einschätzungen, daß Noten und Prüfungen im wesentlichen die mißerfolgsmeidende Komponente der Leistungsmotivation aktivieren, über die der Streit noch nicht beigelegt ist, ob sie eine hemmende oder eine zusätzlich aktivierende Komponente darstellt. Nimmt man an, daß die Anregung von Furcht vor Mißerfolg die resultierende Gesamtmotivation eher verringert, dann müßte die durch die Prüfungen erzeugte Furcht vor Mißerfolg (AHMANN & GLOCK 1967; NUNNALLY 1972) nach der Prüfung wegfallen, und die resultierende Motivation müßte ansteigen. Diese Wirkung aber könnte unabhängig von der differenzierten Rückmeldung stattfinden; die Änderung ist aber als von den personenspezifisch unterschiedlich ausgeprägten Motivtendenzen abhängig zu sehen.

Diese Anmerkung verweist auf ein allgemeines Problem, das abschließend Erwähnung verdient; es ist nicht geklärt, ob durch Noten und Prüfungen überhaupt die Leistungsmotivation als dominante Motivation angesprochen wird, wie dies FISCH (1970), RAYNOR (1968) und viele andere nahelegen, oder ob nicht vielmehr extrinsische Tendenzen im Sinne von ATKINSON (1964) angeregt werden, wie dies die Ergebnisse der Untersuchungen von McKEACHIE (1967) und RAYNOR (1970) belegen. RAYNOR berichtet eine von der Leistungsmotivation unerwartet relativ unabhängige und bedeutsame Wirkung der

erlebten Instrumentalität von Noten auf die Höhe der erzielten Leistung.

Zur Rolle der intrinsischen Motivation

Die abschließenden Bemerkungen des vorangehenden Abschnittes machen deutlich, daß die Leistungsmotivation als eine Form der intrinsischen Motivation gesehen werden kann. Es spricht aber viel dafür, davon eine andere Form der intrinsischen Motivation abzugrenzen, die als Interesse an der spezifischen sachgesetzlichen Auseinandersetzung mit einem besonderen Gegenstand verstanden wird.

Die Arbeiten von ROSENFELD (1966) und FEND (1971) lassen es als naiv erscheinen, Leistungsaspekte und soziale Bewertungen als einem solchen Sachbezug nicht zugehörig (extrinsisch) auszuklammern. Insofern stellt eine Einordnung der Handlungsplanung als intrinsische Motivation (die außerdem rein begrifflich bleibt und, wie Kapitel III.I.D. zeigt, nicht eine eigentliche theoretische Integration ist) keinen Widerspruch zum Versuch dar, auch aus dem Modell der Leistungsmotivationsforschung Erkenntnisse zu gewinnen. Wie schon als theoretischer Hintergrund ist auch in bezug auf Änderungen durch die Noten die Theorie der intrinsischen Motivation selbst wenig ergiebig. Allein zwei empirische Untersuchungen aus ihrem Umkreis versprechen anwendbare Ergebnisse.

Die Untersuchung von MAEHR & STALLINGS (1972) stellt eine Verbindung zum Problem der Bewertung her. Sie unterscheidet zwischen sachlicher und äußerer Bewertung (internal/external evaluation). Sachlich orientierte Bewertung liegt dann vor, wenn die Person allein aus Interesse an der Sache arbeitet und nur die Richtigkeit von Lösungen mitgeteilt bekommt; äußere Bewertung liegt vor, wenn die Aufgabe als Test der Fähigkeit angekündigt wird und die Ergebnisse an den Lehrer wei-

tergegeben werden sollen. Unter diesen Bedingungen hängt die Motivation als nachträgliche Bereitschaft zur freiwilligen Beschäftigung mit den Aufgaben von der Art der Aufgaben ab. Nur bei schwierigen Aufgaben ist bei sachlich orientierter Bewertung die Bereitschaft zur Weiterarbeit größer, bei leichteren Aufgaben dagegen bei äußerer Bewertung. Die Ergebnisse können aber in einer Wiederholungsuntersuchung so nicht vollständig bestätigt werden. Dort, wo im Gegensatz zur ersten Untersuchung nicht im wesentlichen Akademikerkinder, sondern Arbeiterkinder an den Untersuchungen teilnehmen, ist der Effekt nur bei Jungen mit hoher Leistungsmotivation nachweisbar.

Bei der Übertragung der Ergebnisse ist zu bedenken, daß sowohl von den handelnden Personen als auch von der Art des Motivationsgegenstands her deutliche Unterschiede zur vorliegenden Situation bestehen. Deutet man die vorliegende Situation eher als äußere Bewertung, wie dies im Vergleich mit dem entsprechenden Verfahren im Experiment naheliegt, dann ist eine Verringerung der sich in Handlungsplanung äußernden sachgerichteten Motivation, die wohl eher dem Interesse an schwierigen Aufgaben entspricht, zu erwarten. Außerdem bestätigt das Ergebnis die schon aus der Leistungsmotivationsforschung begründete Vermutung komplexer Interaktionen mit Persönlichkeitsbedingungen.

Eine zweite Untersuchung aus dem Begriffsbereich der intrinsischen Motivation (DECI 1972) nimmt eine zusätzliche Differenzierung vor, indem auf der Seite extrinsischer Belohnung zwischen Geld und verbalem Lob unterschieden wird. Die freiwillige Bereitschaft zum Weiterarbeiten steigt nach verbalem Lob, sinkt aber nach Geld. Das wird attributionstheoretisch dadurch erklärt, daß das Geld den erlebten Schwerpunkt der Kontrolle eigenen Verhaltens (locus of cause)

nach außen verschiebt, während verbales Lob als Information über das Erleben eigener Kompetenz wirkt. Die Wirkung der Noten in Prüfungen als eine Form des verbalen Lobs hängt davon ab, ob die Prüflinge die Noten als Urteile über ihre Kompetenz erleben können, was auf Grund der Einstellungsuntersuchung von TIMMERBRINK (1971) bezweifelt werden kann. Die Abhängigkeit der Wirkung von der sozialen und intrapersonalen Dynamik der Situation wird daraus deutlich, daß in der Untersuchung von DECI das berichtete Ergebnis nur für männliche Versuchspersonen gilt, was auf eine Interaktion mit dem Geschlecht des Versuchsleiters zurückgeführt wird.

Mit beiden Untersuchungen sind über den Begriff der intrinsischen Motivation positive Wirkungen der Noten auf die Handlungsplanung eher auszuschließen. Für die von den Normwerten geforderte differenzierte Wirkung als positiver Zusammenhang von Höhe der Note und Stärke der Motivation finden sich keine Hinweise.

Einfluß von Erfolg und Mißerfolg, Lob und Tadel

Beide Themenstellungen einer Sammlung empirischer Untersuchungen, die in relativer Unabhängigkeit von der Leistungsmotivationsforschung sich entwickelt haben, beschäftigt im wesentlichen der Einfluß der Richtung der Rückmeldung auf die folgende Leistung. Sollen die Ergebnisse auf die vorliegende Situation übertragen werden, ist von der Leistung auf die Handlungsplanung rückzuschließen.

Aus dem Bereich von experimentell induziertem Erfolg und Mißerfolg fällt ein Muster von Ergebnissen auf, das auch schon bei der Änderung des Selbstkonzepts ähnlich vorzufinden war: Der Einfluß der Rückmeldung auf die Leistung wird durch das Selbstwertgefühl der Person vermittelt. Bei Erfolg ist bei positivem Selbstwertgefühl eine größere Leistungszunahme

festzustellen; bei Mißerfolg ist bei negativem Selbstwertgefühl eine größere Leistungsabnahme festzustellen (SHRAUGHER & ROSENBERG 1970; SILVERMAN 1964). Die Arbeit von SHRAUGHER & ROSENBERG bestätigt darüber hinaus, daß insgesamt gesehen Erfolg sich eher positiv auf die folgende Leistung auswirkt. (Der Effekt wird hier, was die Übertragung erschwert, an einer anderen als der Leistung untersucht, für die die Rückmeldung erfolgt.) Doch ist selbst dieser Zusammenhang nicht unter allen Umständen bestätigt. So finden MARECEK & METTEE (1972), daß die moderierende Wirkung des Selbstwertgefühls nur dann gilt, wenn das Selbstwertgefühl eine große subjektive Sicherheit aufweist.

Die Ergebnisse lassen wiederum die differenzierte Wirkung des Ausmaßes von Erfolg und Mißerfolg nicht erkennen. Das ist aber für die vorliegende Situation wesentlich, da bei der Eigenart der Notenverteilung mit wenig klarem Mißerfolg zu rechnen ist. Aus diesem Grunde sind auch Untersuchungsergebnisse aus dem Bereich von "Lob und Tadel", wie sie für eine schon Jahrzehnte zählende Tradition von KENNEDY & WILLCUTT (1964) zusammengefaßt werden, eine nur geringe Erkenntnishilfe. Die Aufarbeitung der Ergebnisse bestätigt insgesamt die von SHRAUGHER & ROSENBERG (1970) gefundenen über die eher positive Bedeutung des Erfolgs: Lob hat eher positive Auswirkungen auf die Leistung als Tadel. Allerdings erweist sich diese Beziehung als stark situationsabhängig.

Einfluß diagnostischer Information

Untersuchungen über den Einfluß der Häufigkeit der Testanwendung auf die Motivation bzw. die Leistung, die in Kapitel II erwähnt wurden, sind auf diese besondere Situation nicht übertragbar und werden dementsprechend nicht ausgewertet. Die Untersuchungen einmaliger Rückmeldung zerfallen in zwei Gruppen; nur eine davon

mit einigen wenigen Arbeiten verwendet Motivationsvariablen als abhängige Größen, während die andere den Einfluß auf Leistungen untersucht und nur über den fragwürdigen Rückschluß von der Leistung auf die Motivation brauchbar ist.

Die direkte Wirkung der Rückmeldung von Testergebnissen auf die Motivation wird nur von einer Untersuchung nachgewiesen, die bei öffentlicher Erwähnung eines guten Testergebnisses höhere Berufspläne, insbesondere bei Studenten aus unterprivilegierter sozialer Herkunft, feststellt (THISTLETHWAITE 1959). Es ist aber zu bezweifeln, ob die Noten im Rahmen von Prüfungen einen Öffentlichkeitscharakter ähnlicher Art besitzen. WESTBROOK (1967) findet bei Sekundarschülern zwar eine Änderung des Selbstkonzepts eigener Fähigkeit (s. Teil 1), die den mitgeteilten Testergebnissen folgt, doch keine entsprechende Änderung der Berufsziele. Nur mit Bedenken dieser Gruppe zuzurechnen ist die Untersuchung von COMPTON (1941), in der Studenten die Wirkung der Rückmeldung von Testergebnissen nach mehreren Monaten selbst einschätzen sollen. Dabei berichten die Studenten, die Erfolg oder Mißerfolg hatten, mehr Zielerhöhungen als Studenten, deren Testergebnisse den Erwartungen entsprachen (ein Ergebnis, das im Sinne der ursprünglichen Normwerte einer kurvilinearen Beziehung ausfällt).

In die zweite Gruppe von Untersuchungen gehören die beiden schon im Hinblick auf den Beitrag der Leistungsmotivationsforschung ausgewerteten Arbeiten von FLOOK & SAGGAR (1968) und FLOOK & ROBINSON (1972). Sie beschränken sich aber auf den Nachweis, daß die Rückmeldung von Testergebnissen insgesamt gesehen einen positiven Einfluß auf die Leistung hat; auch das ist nur in einer der beiden Untersuchungen belegt. Seit den Arbeiten von ROSS (1933), die die aus Laborexperimenten an einzelnen Lernakten bekannte positive

Wirkung rückmeldender Information auf komplexe Lernleistungen unter schulischen Bedingungen nicht bestätigen konnten, ist auf diesem Forschungsgebiet der Rückmeldung über den Erfolg von Lernsequenzen wenig neue Erkenntnis zusammengetragen worden (zusammenfassend: KIRKLAND 1972; ROSS & STANLEY 1954). MEANS & MEANS (1971) berichten nach einer manipulierten Rückmeldung von Testergebnissen, die das Bestehen eines Kurses vorhersagen sollen, eine Interaktion mit der Leistungsfähigkeit der Studenten: Gute Studenten zeigen nach negativer Rückmeldung, schlechte Studenten nach positiver Rückmeldung eine höhere Leistung.

Auch die hier zusammengetragenen Erkenntnisse sprechen insgesamt nicht dafür, daß eine Auswirkung der Noten auf die Motivation im Sinne der Normwerte erfolgen wird.

Einfluß der Noten

Wie schon im entsprechenden Abschnitt des Teils 1 fehlen systematische Untersuchungen. Als weiteres Problem kommt hinzu, daß zumeist die mögliche Wirkung auf die Motivation aus Wirkungen auf Leistungen erschlossen werden muß.

Die Fallstudie von HETZER (1933) berichtet sehr verschiedenartige Reaktionen der Schüler auf Zeugnisse, die ihren Beobachtungen nach u.a. mit dem Geschlecht der Schüler zusammenhängen. Insgesamt gesehen scheinen gute Zensuren eher eine motivierende Wirkung zu haben. Dem entspricht der Nachweis der Verstärkung meidender Tendenzen durch schulischen Mißerfolg (ALEXANDER 1968; HÖHN 1967). BIRNEY (1964) betont, daß, bezogen auf über längere Zeit erstreckte Ziele und deren Zusammenhang mit Noten, Untersuchungen fehlen. Eine eigene Fragebogenuntersuchung läßt ihn zu dem Urteil kommen, daß die motivierende Wirkung von Noten sehr fraglich sei, weil die Studenten kaum Ver-

trauen in Noten haben und das sachliche Interesse an einem Kurs die wirksamere Motivation sei.

Untersuchungen über die Auswirkung der Notengebung auf Leistungen gehen häufig so vor, daß an den Benotungen oder den Notensystemen Änderungen vorgenommen werden, um auf dem Hintergrund der "normalen" Situation durch gezielte Änderung bestimmter Elemente einen Wirkungsnachweis zu führen. Bei diesem Vorgehen verlieren solche Felduntersuchungen aber das, was sie manipulierenden Experimenten gegenüber voraushaben: Auch sie stellen jetzt neuartige Situationen dar, deren Neuartigkeit allein manchen Unterschied zu den "herkömmlichen" Noten bewirken kann (Hawthorne-Effekt).

Zwei Untersuchungen gelten der Änderung des gesamten Notensystems. Nach FAY (1937) bringt die Einführung eines differenzierten Systems gegenüber einem wenig differenzierten dreistufigen System eine Leistungssteigerung bei guten und schlechten Studenten mit sich. BAKER & DOYLE (1958) führen im letzten Jahr vor dem Übergang zur High-School an Stelle des in der Schule üblichen individualisierten Benotungsprogramms ein vergleichsorientiertes Benotungsprogramm ein. Die Leistung im folgenden Jahr der High-School unterscheidet sich nicht von derjenigen des Schülerjahrgangs des vorhergehenden Jahres, der durchgehend individualisierte Noten erhielt.

Andere Untersuchungen manipulieren die rückgemeldeten Noten, um den Einfluß auf die Leistung zu ermitteln. BEYER (1971) erhöht willkürlich den zu Anfang des College-Besuchs mitgeteilten GPA (grade point average) um eine Standardabweichung. Weder in der Leistung nach einem Jahr noch in Daten über die Studienaktivitäten zeigen sich irgendwelche Unterschiede im Vergleich mit einer Studentengruppe mit dem richtigen GPA. (Im übrigen ist an dieser Stelle die Grenze überschritten, an der Manipulation durch Erkenntnisinter-

esse sich legitimieren läßt.) PICKUP & ANTHONY (1968) gestalten die Erfolgsrückmeldung an Schüler in Abhängigkeit von deren Erwartung. Gute Schüler zeigen dann mehr Leistungen, wenn die Rückmeldung unter ihrer Erwartung bleibt; schlechte Schüler zeigen dann mehr Leistungen, wenn die Rückmeldung ihre Erwartung übertrifft, obwohl die Erwartung selbst schon über den auf Grund der vergangenen Leistungen zu erwartenden Leistungen liegt. Ein ähnliches Ergebnis wurde im letzten Abschnitt bei der Rückmeldung von Testergebnissen berichtet (MEANS & MEANS 1971).

Einen nur indirekten Hinweis auf die vermutlich geringfügige motivierende Wirkung einfacher Mitteilung von Noten als Zahlenwerten gibt die sorgfältig geplante Untersuchung von PAGE (1958), die durch das Hinzufügen von Kommentaren zu den wie üblich vergebenen Noten eine deutliche Leistungssteigerung, insbesondere bei schlechten Schülern, bewirkt. Eine Replikation von FITTKAU & LANGER (1974) allerdings kann eine entsprechende Leistungsveränderung nicht feststellen, was umgekehrt noch kein Hinweis auf die motivierende Wirkung der als Zahlen mitgeteilten Noten ist.

Situationsbezogene Integration der Ergebnisse

An den entsprechenden Stellen der vorangehenden Abschnitte wurde jeweils schon der Versuch unternommen, kurz die Übertragbarkeit der Erkenntnisse anzusprechen. Es wird aus diesen Anmerkungen deutlich, daß selten die Situationen der Erkenntnisgewinnung mit der vorliegenden vergleichbar sind. Doch die Vielzahl aller Arbeiten, die kaum einmal eine klare Wirkung im Sinne der Normwerte nachweisen, läßt das zusammenfassende Urteil zu, daß normentsprechende Wirkungen der Noten auf die Handlungsplanung nicht in bedeutsamem Ausmaß zu erwarten sind.

Einzig die besondere Art der Notenvergabe (s. Kapitel III.I.B.) könnte durch den Verzicht auf gravierenden Mißerfolg und entsprechende Betonung von Erfolg diese Situation vor denjenigen auszeichnen, an denen die berichteten Ergebnisse gewonnen wurden. Da gerade der Mißerfolg zu keinen oder sogar atypischen Reaktionen Anlaß gibt, kann sein Wegfall eine normgerechte Wirkung ermöglichen.

Gegen eine solche Annahme sprechen zwei weitere Merkmale der besonderen Situation: Die Motivation als Prozeß emotionaler Wertung ist in der Form der Handlungsplanung analytisch in einen emotionalen und einen kognitiv-rationalen Prozeß aufspaltbar. Der emotionale Prozeß beinhaltet u.a. die Mitwirkung persönlichkeitsspezifischer Bedingungen. Es ist anzunehmen, daß in einer Situation, die leistungsbezogen emotional in großem Umfang aktiviert, auch der Einfluß von Persönlichkeitsbedingungen auf Prozesse, die mit dieser leistungsbezogenen Aktivierung eine Verbindung aufweisen, zunimmt. Eine Zunahme von Persönlichkeitseinflüssen auf die Verarbeitung der Erfahrung wird eine Abweichung dieser Verarbeitung von den vorgegebenen Normwerten zur Folge haben. Die rationale Komponente der Handlungsplanung ist in dem sachlichen Zusammenhang verschachtelter Ziele zu sehen: In einer hierarchischen Struktur von Handlungszielen bestimmen die allgemeineren die konkreteren. Die beabsichtigte Intensität der Weiterarbeit stellt ein konkretes Ziel in der Hierarchie von Lebens- und Berufszielen dar. Sie hat sich in allgemeine Lebens- und Berufsziele einzuordnen. Nur in dem Maße, in dem diese selbst Adressaten von Leistungsrückmeldungen in Prüfungen sind, wäre sachlich eine davon abgeleitete Änderung der konkreten Ziele notwendig. Gerade im VD aber sind vermutlich die allgemeineren Ziele von der Fachstruktur des VD völlig losgelöst, so daß die rationale Be-

gründung der Weiterarbeit in den Fächern des VD sich aus derem sachlichen Beitrag zu den Zielen des HD-Studiums usw. ergibt, was sie von den konkreten Leistungen in den Fächern des VD relativ unabhängig macht. Im HD ist zwar eine größere Identität von Fächern und folgender Lebens- und Berufsperspektive vorhanden, doch ist auch hier noch ein Bruch besonders zur praktischen, aber auch zur wissenschaftlichen Tätigkeit wahrscheinlich.

Zusammenfassung (III.II.A.2.)

Die Auswertung von vorliegenden Erkenntnissen ist auf die Handlungsplanung zugeschnitten. Dabei wird die Übertragung dadurch belastet, daß Handlungsplanung nicht mit den Gegenständen von Theorien und empirischen Untersuchungen zur Motivation identisch ist.

In der Theorie der Leistungsmotivation ist wenig positive Evidenz zu finden. Deren Ergebnisse machen aber darauf aufmerksam, daß personal gebundene Unterschiede in einen möglichen Zusammenhang bedeutsam eingreifen können. Zwei Untersuchungen aus dem Begriffsbereich der von der Leistungsmotivation abgesetzten intrinsischen Motivation machen wiederum komplexe Interaktionen deutlich, während Hinweise auf eine normgerechte Wirkung der Rückmeldung auf die Handlungsplanung fehlen. Auch Untersuchungen über den Einfluß von Erfolg und Mißerfolg bzw. von Lob und Tadel auf die folgende Leistung sprechen, wenn man von der Leistung auf die Motivation zurückschließt, nicht klar für die geforderte Wirkung. Ähnliches ist an den Ergebnissen für die Rückmeldung von diagnostischen Informationen festzustellen. Untersuchungen an den Noten selbst sind wie beim Selbstkonzept selten und lassen zwar Einflüsse der Noten erkennen, die aber komplexeren Mustern folgen.

Zusammenfassend ist nicht mit einer Wirkung auf die Handlungsplanung im Sinne der Normwerte zu rechnen. Von den konkreten Eigenschaften der Situation spricht das Ausbleiben von gravierendem Mißerfolg für eine normgerechte Wirkung, während die vermutliche emotionale Aktivierung eine Zunahme des verfälschenden Persönlichkeitseinflusses nahelegt und die rationale Komponente der Handlungsplanung durch die Funktionalisierung der Beschäftigung mit den Prüfungsfächern im Hinblick auf weitererstreckte Ziele die Motivation von den gegenwärtigen Leistungen ablöst.

III.II.B.: Empirische Analyse der Wirkungen

Übersicht: Die Wirkungsanalyse bedient sich verschiedener Methoden des Zugangs zu den Daten, deren gemeinsamer Nenner ist, daß die Noten (und Prüfungen) als unabhängige Variablen und die thematischen Fragen als abhängige Variablen betrachtet werden. In zunehmendem Maß wird versucht, den globalen Prozeß der Wirkung auf dem Hintergrund von ganz bestimmten Bedingungen zu betrachten und deren Anteil an dem Einfluß zu bestimmen. Der Auswertung vorangestellt ist der für mehrere Auswertungsschritte gültige Versuch einer Begründung der gewählten Methodender Datenauswertung.

Eine direkte Überprüfung der Wirkung der Gesamtnoten und der Noten in den einzelnen mündlichen Prüfungen auf Selbstkonzept und Motivation beabsichtigt Teil 1 der Ergebnisdarstellung. Die beiden Prüfungsarten werden, wie auch in allen folgenden Schritten der Auswertung, als getrennte Datensätze behandelt.

Die beiden folgenden Teile untersuchen die Wirkung unter bestimmten Bedingungen. Eine erste Bedingung stellt die Prüfungsart (VD und HD) dar; sie wird bereits im ersten Teil einbezogen. Im Teil 2 werden darüber hinaus Geschlechtsunterschiede und Unterschiede der gewählten Prüfungsordnung (nur für das VD) als

weitere mögliche Bedingungen der Wirkungen kontrolliert. Eine "Mitwirkung" dieser Bedingungen würde der Annahme universeller Wirkungen, die mit den Normwerten impliziert wird, widersprechen.

Ein zweiter Ansatz der Analyse von konkreten Bedingungen der Wirkung gilt den in Teil 1 berichteten unterschiedlichen Ergebnissen für einzelne "Fach-Prüfer-Kombinationen". Es wird in Teil 3 geprüft, ob diese Unterschiede mit bestimmten Eigenschaften der "Fach-Prüfer-Kombinationen", wie sie aus den gemittelten Antworten zu anderen Fragen hervorgehen, zusammenhängen.

Nicht alle Änderungen über die Prüfung hinweg sind als Abhängigkeiten von den erteilten Noten zu sehen. Teil 4 soll aus den Veränderungen der Kontingenzen zwischen den thematischen Fragestellungen innerhalb der beiden Fragebögen ergänzende Hinweise auf den Einfluß der Prüfungen, der über die Noten hinausgeht, gewinnen.

Im Teil 5 wird in entsprechender Absicht die Änderung der Parameter der Antwortverteilungen über die Prüfung hinweg untersucht.

Mit der Auswertungsmethode der Faktorenanalyse werden im abschließenden Teil 6 die auf der Ebene "Über-alle-Fächer" berechneten Korrelationen zwischen den Antworten auf solche Fragen, die sich auf die einzelnen mündlichen Prüfungen beziehen, ausgewertet, um aus der Einbettung der thematischen Fragen in das Fragengesamt weitere Aufschlüsse über Bedingungen der Wirkungen zu erhalten.

Methodische Grundlagen der Wirkungsanalyse

Die zentralen Fragestellungen der empirischen Auswertung sind die nach Bestehen und Ausmaß von Wirkungen der Noten auf die abhängigen Variablen. Zur Beantwortung der Fragestellungen würde eine Auswertung

des Zusammenhangs von Noten und Antworten im PFF nicht ausreichen. Ein solches Vorgehen der Auswertung würde annehmen, daß der Zusammenhang nicht schon "vor der Prüfung" bestanden hat. Das aber ist durchaus möglich, indem beispielsweise ein höheres Selbstkonzept eigener Fähigkeit vor der Prüfung sich in einem leistungsbezogenen Verhalten des Prüflings auswirken kann, das mit einer besseren Note bewertet wird. Bliebe das Selbstkonzept über die Prüfung hinweg unverändert, wäre zwischen Noten und der Antwort nach der Prüfung ein entsprechender Zusammenhang feststellbar, der aber gar nicht auf die Wirkung der Noten zurückgeht. (Andere inhaltlich differenziertere Modelle sind denkbar; an dieser Stelle geht es allein um das formale Schema solcher "Schein"kausalität.) Dem Bedenken wird durch die Anlage der Untersuchung als prae-post-design Rechnung getragen, die eine Isolierung der Zusammenhänge zwischen Noten und Antworten, die nach der Prüfung auf den Einfluß der Noten zurückgeführt werden können, von solchen Zusammenhängen erlaubt, die vor der Erteilung der Noten sich konstituieren (damit ist nicht gesichert, daß nicht mit den Noten konfundierte Wirkgrößen wie z.B. der Verlauf der Prüfung selbst die eigentliche wirkende Ursache an Stelle der Noten sind).

Statistische Probleme einer solchen Isolierung der Zusammenhänge entstehen durch die komplexe Struktur der Daten und die Eigenart der Notenverteilungen. Es stehen verschiedene Auswertungsmöglichkeiten zur Verfügung. Eine erste naheliegende Möglichkeit und in der bisherigen Selbstkonzeptforschung häufig verwendete Methode berechnet Differenzwerte zwischen der ersten und der zweiten Messung der abhängigen Variablen und verwendet diese als neue Meßwerte. Korrelationen der Noten mit diesen Änderungswerten wären ein Maß des von den Noten ausgehenden Einflusses. Gegen ein solches Vorgehen haben CRONBACH & FURBY (1970),

u.a. im Hinblick auf die Diskrepanzwerte der Selbstkonzeptforschung, Bedenken geltend gemacht, indem sie darauf hinweisen, daß von einem meßtheoretischen Standpunkt aus Einwände erhoben werden müssen. Die Fehlerkomponenten in den beiden Messungen derselben Variablen können nicht als voneinander unabhängig betrachtet werden. Die Schätzung des Änderungswertes aus der Bildung von Differenzen zwischen den beiden Messungen wird deshalb nicht von einem systematischen Fehler frei sein, der nur durch eine von beiden Autoren in Anlehnung an TUCKER entwickelte Regressionsschätzung ausgeglichen werden kann, die die Zuverlässigkeit der beiden Messungen mit einbezieht. Daten darüber aber fehlen in dieser Untersuchung. Für den vorliegenden Zweck ist der Hinweis von CRONBACH & FURBY nützlich, daß in vielen Fällen, in denen solche Differenzwerte bestimmt werden, die Berechnung von partiellen Korrelationen denselben Zweck erfülle und methodisch einwandfrei sei. Sie stellen als weitere Methode die Kovarianzanalyse zur Diskussion, von deren Verwendung sie aber abraten, wenn die Personen den Bedingungen nicht zufällig zugeordnet werden.

Die Auswahl der Methoden hängt davon ab, wie weit die Daten die jeweils geforderten Voraussetzungen erfüllen. Als Verfahren kommt neben der partiellen Korrelation und der Kovarianzanalyse eine parameterfreie Auswertung in Frage.

Die partielle Korrelation bestimmt den Zusammenhang von zwei Variablen unter Ausschaltung (Konstanthaltung) des vermittelnden Einflusses einer dritten (GUILFORD 1973). Bei dem Zusammenhang zwischen Noten und Selbstkonzept nach der Prüfung wird das Selbstkonzept vor der Prüfung konstant gehalten und damit sein Zusammenhang mit den Noten auspartialisiert. Sie setzt voraus, daß lineare Regressionen zwischen den Variablen bestehen; ferner müssen die Variablen intervall-

skaliert sein. Die Signifikanzprüfung setzt Normalverteilung voraus.

Ein weiteres parametrisches Verfahren stellt die Kovarianzanalyse dar. Sie ermöglicht es, auch das Zusammenwirken von mehreren Einflußquellen und deren Interaktionen zu erfassen. An die abhängigen Variablen stellt sie eine Reihe von Forderungen, wie Gleichheit der Varianzen, Intervallskalierung usw. Die unabhängige(n) Variable(n) allerdings brauchen solche Forderungen nicht zu erfüllen, sondern können nominalskaliert vorliegen. Für die Einbeziehung von Kovariaten, die auspartialisiert werden, ist außerdem zu fordern, daß Interaktionen von Faktoren und Kovariaten nicht bestehen (NIE et al. 1975).

Verfahren, die keine Verteilungsvoraussetzungen machen, sind zur Auspartialisierung nur für Rangdaten vorhanden (LIENERT 1973). Rangdaten aber würden infolge der Eigenarten der Notenverteilungen einen hohen Anteil an Bindungen erzeugen, deren Korrektur nicht ohne Probleme ist. Auf der Ebene von Nominaldaten besteht nur die Möglichkeit, Differenzwerte zwischen den beiden Antworten zu berechnen und als Ausgangsdaten zu verwenden, um einen ursächlichen Zusammenhang mit den Noten über eine Vierfelder-Tafel (Fisher-Yates-Test) festzustellen.

Für die abhängigen Variablen kann angenommen werden, daß sie die Verteilungsvoraussetzungen erfüllen (s. Exkurs I und Kapitel III.I.E.). Bei den Noten dagegen bestehen Zweifel an der Intervallskaliertheit; auf jeden Fall aber ist eine Normalverteilung nicht gegeben (s. Kapitel III.I.B.). Im Hinblick auf die konkreten Eigenschaften der Daten verstärkt sich der Eindruck, daß keine der Auswertungsmethoden für die Analyse der globalen Wirkung unproblematisch ist.

Die partielle Korrelation setzt eine Normalverteilung der Noten voraus, die nicht erfüllt ist.

Die Kovarianzanalyse stellt an die Noten zwar keine Anforderungen, setzt aber zufällige Zuordnung der Personen zu den Faktoren voraus, die dem Charakter der Felderhebung widerspricht.

Die parameterfreie Auswertung macht auf seiten der verwendeten Variablen keine Voraussetzungen, führt aber die problematische Berechnung von Änderungswerten wieder ein.

Zieht man die Ausschöpfung der in den Variablen enthaltenen Information in Betracht, dann entsteht für die drei Verfahren durch die verfahrenstechnische Konkretisierung ein Gefälle: Die partielle Korrelation nutzt die gegebene Information am besten. Die Kovarianzanalyse macht eine Reduktion der Noten als Faktoren erforderlich, die am besten in zwei Klassen (Notenstufe "1", Notenstufe "schlechter als 1") erfolgt. Dadurch geht unter vielen Auswertungsbedingungen Information verloren. Die parameterfreie Auswertung auf Nominalskalenniveau läßt die gleiche Dichotomisierung bei den Noten sinnvoll erscheinen, zu der aber ein zusätzlicher Verlust an Information durch Dichotomisierung der Differenzen hinzutritt.

Die Darstellung macht deutlich, daß es keine ohne Einschränkung geeignete Methode gibt. An Stelle einer generellen Entscheidung soll versucht werden, die Methoden parallel anzuwenden. Bestimmte Fragestellungen nach dem Einfluß von Bedingungen als weiteren Wirkungen oder nach dem unterschiedlichen Einfluß unter verschiedenen Bedingungen sind aber nur mit den parametrischen Methoden auszuwerten.

Ein zusätzliches allgemeines methodisches Problem betrifft die Auswertung der Motivationsfragen. Ursprünglich wurden als Normwerte für diese Fragen kurvilineare Beziehungen festgelegt, die unter Berücksichtigung der einseitigen Verteilung der Noten auf die Norm eines positiv gerichteten Einflusses reduziert

wurden. Für die Auswertung ist genauer zu bestimmen, was mit einer positiven Beziehung gemeint ist. Beim vergleichsorientierten Selbstkonzept eigener Fähigkeit ist offensichtlich, daß in den Fächern die Noten mit den Selbstkonzepten der einzelnen Personen einen positiven Zusammenhang aufweisen sollten. Die Messung der Motivation aber erfolgt nicht vergleichsorientiert, sondern nach Art eines absoluten Urteils. Ein positiver Zusammenhang könnte sich sinnvoll darauf beschränken, daß ein Prüfling in dem Fach (oder den Fächern), in denen er die besseren Noten hat, eine intensivere Weiterarbeit als vor der Prüfung plant und in den übrigen Fächern eine weniger intensive. Im Hinblick auf diese Formulierung wäre eine Profilauswertung über die Einzelfächer hinweg angebracht. Da jedoch weder die Noten miteinander hohe Korrelationen aufweisen, noch die abhängigen Fragen nach der Motivation, kann die methodisch aufwendige und mit neuen Voraussetzungen belastete Profilauswertung durch eine Berechnung über die Personen und über die Prüfungen ("Über-alle-Fächer") hinweg ersetzt werden. Damit kann die Auswertung derjenigen des Selbstkonzepts eigener Fähigkeit strukturgleich erfolgen.

III.II.B.1.: Globale Wirkung der Noten

Die Analyse der Wirkung ist durch die Eigenschaften der Notenverteilungen begründet nicht auf allen vorgeschlagenen Auswertungsebenen möglich. Im HD kann sie für die Einzelfächer nicht in den einzelnen "Fach-Prüfer-Kombinationen" und für die Gesamtwerte nur in der "Weiten Stichprobe" durchgeführt werden.

Wirkung der Noten in den "Fach-Prüfer-Kombinationen"

Auch im VD müssen zwei "Fach-Prüfer-Kombinationen" von der Auswertung ausgeschlossen werden, weil dort

die von der Notenstufe "1" abweichenden Bewertungen so selten sind, daß eine Auswertung nicht sinnvoll erscheint. Es werden solche Kombinationen nicht ausgewertet, bei denen weniger als fünf vonder Notenstufe "1" abweichende Noten festzustellen sind. Nach Tabelle: III.I. - 1 sind das die Kombinationen Allg.Psychol./14 und Pers.Psychol./02.

Tabelle: III.II. - 1

Wirkung der Noten in den einzelnen "Fach-Prüfer-Kombinationen" des VD

Fach-Prüfer-Kombinationen	Method./02	Ausdr./09	Allg.Psychol./12	Entw.Psychol./05	Pers.Psychol./05	Physiol./08	Philos./01	Philos./03	Humang./07
Zahl der Prüflinge	16	18	33	42	30	40	10	21	29
PVF/PFF-SK.FÄCHER									
Richtung der Änderung der Korrelat. Note/PFF gegenüber Note/PVF	+	+	+	+	+	+	+	-	+
partielle Korrel. Note/PFF (PVF auspart.)	+.63	+.22	+.41	+.47	+.50	+.25	+.78	-.07	+.15
Wahrscheinlichkeit zufäll.Abw. von $r_{part.} = .00$	$<.05$	n.s.	$<.05$	$<.01$	$<.01$	n.s.	$<.05$	n.s.	n.s.
Wahrscheinlichkeit zufäll.Zusammenh. Noten/Differenzen PVF - PFF	n.s.	n.s.	n.s.	$<.05$	n.s.	n.s.	n.s.	n.s.	n.s.

Fortsetzung der Tabelle s.S. 232

Tabelle: III.II. - 1 (Fortsetzung)

Wirkung der Noten in den einzelnen "Fach-Prüfer-Kombinationen" des VD

Fach-Prüfer-Kombinationen	Method./02	Ausdr./09	Allg.Psychol./12	Entw.Psychol./05	Pers.Psychol./05	Physiol./08	Philos./01	Philos./03	Humang./07
Zahl der Prüflinge	16	18	33	42	30	40	10	21	29
PVF/PFF-MOT.W.A.									
Richtung der Änderung der Korrelat. Note/PFF gegenüber Note/PVF	+	+	-	+	+	+	-	-	+
partielle Korrel. Note/PFF (PVF auspart.)	.31	.43	-.11	.16	-.07	.13	-.06	-.06	.37
Wahrscheinlichkeit zufäll.Abw. von $r_{part.} = .00$	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Wahrscheinlichkeit zufäll.Zusammenh. Noten/Differenzen PVF - PFF	n.s.	n.s.	n.s.	$<.05$	n.s.	n.s.	n.s.	n.s.	n.s.
PVF/PFF-MOT.BEL.									
Richtung der Änderung der Korrelat. Note/PFF gegenüber Note/PVF	+	+	+	+	-	+	+	-	+
partielle Korrel. Note/PFF (PVF auspart.)	.27	.19	.10	-.10	-.21	.32	.51	-.07	.42
Wahrscheinlichkeit zufäll.Abw. von $r_{part.} = .00$	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	$<.05$
Wahrscheinlichkeit zufäll.Zusammenh. Noten/Differenzen PVF - PFF	n.s.	n.s.	n.s.	n.s.	$<.05$	n.s.	n.s.	n.s.	$<.01$

Methodisches Vorgehen: Für die drei thematischen Fragen sind jeweils vier Zeilen mit Informationen für jede "Fach-Prüfer-Kombination" vorgesehen. In der ersten Zeile ist die Richtung der Änderung der Korrelation mit den Noten angegeben. Ein positives Vorzeichen markiert eine Änderung im Sinne der Normwerte, d.h. also: Die Beziehung zwischen Noten und thematischen Fragen ändert sich in der Richtung, daß bessere Noten mit größerer Motivation und höherem Selbstkonzept einhergehen. Die zweite Zeile gibt die Berechnung der partiellen Korrelation wieder, deren Signifikanz mit dem üblichen Verfahren (GUILFORD 1973) in der dritten Zeile bestimmt wird. Die partielle Korrelation ist ein Maß für den Zusammenhang zwischen Noten und Antworten im PFF, der nicht auf einen Zusammenhang zwischen Noten und entsprechenden Antworten im PVF zurückgeführt werden kann. Auch die Richtung der partiellen Korrelationen ist so gepolt, daß positive Korrelationen Wirkungen im Sinne der Normwerte anzeigen. Die vierte Zeile beinhaltet das Ergebnis einer parameterfreien Auswertung mit dem Fisher-Yates-Test, der die am Median dichotomisierten Noten und Differenzwerte zwischen PVF und PFF auf einen Zusammenhang hin prüft. Wegen der Problematik der Anwendung der inferenzstatistischen Verfahren sei auf den vorangehenden Abschnitt verwiesen.

Insgesamt gesehen liegen mit einer Ausnahme (Pers.Psychol./05 - MOT.BEL.) alle Zusammenhänge, wenn sie die Grenzen statistischer Signifikanz überschreiten, in der erwarteten Richtung. Die drei thematischen Fragen unterscheiden sich aber deutlich in der Ausgeprägtheit der Ergebnisse.

Für das fachbezogene Selbstkonzept eigener Fähigkeit fallen die partiellen Korrelationen mit einer Ausnahme in der erwarteten Richtung aus. Sie sind sehr unterschiedlich hoch und längst nicht in allen einzel-

nen "Fach-Prüfer-Kombinationen" signifikant. Eine Prüfung auf die statistische Bedeutsamkeit der Unterschiede über alle Korrelationen mit einem von BARTEL (1972) vorgeschlagenen Verfahren kann die Unterschiede nicht gegen den Zufall absichern. Es ist kaum eine Kontrolle möglich, ob die unterschiedliche Höhe der partiellen Korrelationen mehr auf die Rahmenbedingung Prüfer oder mehr auf die Rahmenbedingung Prüfungsfach zurückgeht. Die Möglichkeit, zu einer Abschätzung der wirksamen Rahmenbedingungen zu kommen, ist, wie schon in der Voraussetzungsanalyse dargestellt wurde, beschränkt. Sie ist hier weiter eingegrenzt durch den zusätzlichen Ausschluß von zwei "Fach-Prüfer-Kombinationen". In Teil 3 soll untersucht werden, ob diese Unterschiede in bestimmten Merkmalen der einzelnen "Fach-Prüfer-Kombinationen" begründet sind.

In der parameterfreien Auswertung läßt sich der Zusammenhang nur in einer Kombination bestätigen. Doch geht dort auch mehr als bei den Motivationsfragen durch die Dichotomisierung Information aus der großen Streuung der Antworten verloren.

Für Beliebtheit und Handlungsplanung weisen schon weniger partielle Korrelationen die durch die Normwerte erwartete Richtung auf. Zwar stimmt die Richtung der meisten Korrelationen mit den Normwerten überein; das Ausmaß ist aber deutlich geringer als beim Selbstkonzept eigener Fähigkeit. Nur eine der partiellen Korrelationen überschreitet die Signifikanzgrenze. Die beiden Signifikanzen der parameterfreien Auswertung, die für den Prüfer 05 gelten und für die Beliebtheit eine "negative" Wirkung der Noten anzeigen, sind ohne Zusatzinformation nicht aufzuklären. Einzig die Beeinflussung der Beliebtheit in der Kombination Humang./07 findet eine plausible Erklärung darin, daß eine bedeutsame sachliche Auseinandersetzung mit diesem Fach erst kurz vor der Prüfung zu

beginnen pflegt, so daß die Note sehr eng mit dieser Auseinandersetzung verknüpft ist (alternative Erklärungen sind aber möglich).

Die übereinstimmenden Tendenzen in den einzelnen "Fach-Prüfer-Kombinationen" werden in der Auswertung "Über-alle-Fächer" des folgenden Abschnittes zusammengefaßt.

Wirkung der Noten auf der Ebene "Über-alle-Fächer"

Die Auswertung auf dieser Ebene (s. Exkurs II) betrachtet jede einzelne Prüfung als voneinander unabhängiges Ereignis. Die Voraussetzungsanalyse wies nach, daß die Antworten zu den einzelnen Fächern zwar nicht voneinander unabhängig sind, besonders bei den Motivationsfragen aber so geringe durchgehende Zusammenhänge vorliegen, daß die Anwendung inferenzstatistischer Verfahren nicht allzu problematisch ist. Die Ergebnisse der statistischen Prüfung werden trotzdem mit (?) versehen.

In Kapitel III.I.B.1. wurde die Anwendung einer Notentransformation als Modell der individuellen Verarbeitung der Noten vorgeschlagen. Das soll an dieser Stelle verwirklicht werden, indem die Noten in den "Fach-Prüfer-Kombinationen" auf gleiche Mittelwerte und Streuungen umgerechnet werden. Die Berechnung solcher Notenäquivalente führt bei der Eigenart der Notenverteilungen dazu, daß bestimmte Äquivalente nur bei bestimmten "Fach-Prüfer-Kombinationen" vorkommen, was sicher eine neue Fehlerkomponente in den neuen Werten bedingt.

Tabelle: III.II. - 2

Wirkung der Noten auf der Ebene "Über-alle-Fächer"

	VORDIPLOM			HAUPTDIPLOM		
	PVF/PFF SK.FÄCHER	PVF/PFF MOT.W.A.	PVF/PFF MOT.BEL.	PVF/PFF SK.FÄCHER	PVF/PFF MOT.W.A.	PVF/PFF MOT.BEL.
mittlere part. Korrelation aller Fach-Prüfer-Kombinat.	+.40	+.13	+.10	nicht berechenbar		
Über-alle-Fächer						
part. Korrelation Note/PFF (PVF auspart.)	+.29	+.15	+.17	+.17	+.15	+.24
Wahrscheinlichkeit zufälliger Abweichung von $r_{part.} = .00$	<.01(?)	<.05(?)	<.01(?)	<.05(?)		<.01(?)
part. Korrelation nach Notentransformation mit Wahrscheinlichkeit der zufälligen Abweichung von $r_{part.} = .00$	+.35	+.12	+.11	+.17	+.10	+.14
	<.01(?)	<.05(?)		<.05(?)		
Varianzanalyse mit Auspartialisierung des PVF Wahrscheinlichkeit zuf.Abw.	<.01(?)	<.01(?)	<.05(?)			<.01(?)
parameterfreier Vierfelder-Test Wahrscheinlichkeit zuf.Abw.	<.01(?)		<.05(?)			<.05(?)

Tabelle: III.II. - 2 (Fortsetzung)

Wirkung der Noten auf der Ebene "Über-alle-Fächer"

	VORDIPLOM			HAUPTDIPLOM		
	PVF/PFF SK.FÄCHER	PVF/PFF MOT.W.A.	PVF/PFF MOT.BEL.	PVF/PFF SK.FÄCHER	PVF/PFF MOT.W.A.	PVF/PFF MOT.BEL.
Über-alle-Fächer (nur psychologische Fächer - VD)						
part. Korrelation Note/PFF (PVF auspart.)	.39	.13	.09			
Wahrscheinlichkeit zufälliger Abweichung von $r_{part.} = .00$	$<.01$(?)			nicht berechenbar		
Varianzanalyse mit Auspartialisierung des PVF Wahrscheinlichkeit zuf.Abw.	$<.01$(?)					

Methodisches Vorgehen: Die erste Zeile gibt die über die "Fach-Prüfer-Kombinationen" von Tabelle: III.II. - 1 gemittelten partiellen Korrelationen wieder. Die Mittelung wird wegen der großen Streuung an z-transformierten Werten vorgenommen; die Gewichtung erfolgt an der Zahl der Freiheitsgrade, die N - 4 beträgt. Die beiden folgenden Zeilen geben die Höhe und Signifikanz der über alle Antworten in den Einzelfächern hinweg berechneten partiellen Korrelationen wieder. Die dann folgende Zeile enthält dieselben beiden Werte für die beschriebene Umrechnung der Noten. Positive Korrelationen bedeuten in jedem Fall Einflüsse in Richtung der Normwerte: erhöhtes Selbstkonzept und vergrößerte Motivation bei besseren Noten. In den anschließenden beiden Zeilen folgen die Ergebnisse der Varianzanalyse, die die Antworten zu den entsprechenden Fragen des PVF aus den Antworten zum PFF auspartialisiert, und des Fisher-Yates-Tests, der den Zusammenhang von am Median dichotomisierten Noten mit den entsprechend dichotomisierten Differenzwerten zwischen PVF und PFF prüft. Für das VD folgt zur Kontrolle die Berechnung von partiellen Korrelationen und Varianzanalysen nur für die Antworten, die in den vier psychologischen Fächern der Prüfung gegeben werden. In diesem Fall beträgt die Zahl der Ereignisse (Prüfungen) um 170; sie schwankt etwas wegen einzelner fehlender Antworten. Bei der Auswertung "Über-alle-Fächer" des VD beträgt sie wiederum etwas schwankend um 280, für das HD 145 Prüfungen.

Die Angaben von "n.s." für fehlende Signifikanz sind weggelassen, weil sie wie die konkreten Angaben von Signifikanzgrenzen auch mit einem (?) versehen werden müßten.

Die Ergebnisse für das VD bestätigen für die Wirkung auf das Selbstkonzept eigener Fähigkeit im großen und ganzen die Auswertung der einzelnen "Fach-Prüfer-

Kombinationen". Das Selbstkonzept eigener Fähigkeit wird durch die in den Prüfungen verteilten Noten in statistisch bedeutsamer Weise, aber im Umfang deutlich begrenzt beeinflußt. Die Richtung der Wirkung entspricht der Vorgabe durch die Normwerte. Demgegenüber ist die Wirkung auf Handlungsplanung und Beliebtheit sehr geringfügig und nicht einmal in allen Auswertungsschritten signifikant. Die Reduzierung auf die Antworten in den psychologischen Fächern läßt zudem deutlich werden, daß die geringfügige Wirkung auf die Motivation auf eine komplexe Interaktion mit Fachbedingungen zurückgeht.

Die Ergebnisse für das HD weisen keine entsprechende Wirkung der Noten auf das Selbstkonzept nach; die Wirkung ist nicht einmal auf allen Stufen der statistischen Auswertung "Über-alle-Fächer" signifikant. Statt dessen tritt die Wirkung auf die Beliebtheit deutlicher hervor, wenngleich auch sie im Ausmaß gering bleibt.

Die Transformation der Noten in den "Fach-Prüfer-Kombinationen" auf gleiche Mittelwerte und Streuungen hat nur geringfügige Auswirkungen. Mit aller Vorsicht wird man den Veränderungen der Korrelationen, besonders im VD, entnehmen, daß bei der individuellen Verarbeitung der Noten im Selbstkonzept die Unterschiede von Notenverteilungen einzelner "Fach-Prüfer-Kombinationen" etwas berücksichtigt werden, während die Wirkung auf die Motivation, besonders die Beliebtheit, eher von der absoluten Notenhöhe ausgeht. Aber dieser Deutung liegen nur geringfügige Änderungen der Korrelationen zugrunde.

An dieser Stelle soll eine erste Gegenüberstellung der Ergebnisse und der Normwerte erfolgen. Die Wirkung der Noten auf das Selbstkonzept eigener Leistungsfähigkeit ist nur im VD in statistisch bedeutsamer Weise festzustellen. Das Ausmaß der Wirkung ist

auch dort nicht groß; dabei ist allerdings in Rechnung zu stellen, daß die sehr unterschiedliche Streuung von Noten und Antworten zum Selbstkonzept das Ausmaß des Zusammenhangs verringert. Eine quantitative Festlegung des zu fordernden Ausmaßes an Wirkung erfolgte nicht. Es besteht aber kein Zweifel, daß das empirische Ausmaß eher unter einem normativ sinnvollen Ausmaß liegt. (Auf die Frage, ob das angesichts der Qualität der Noten nicht eine höhere Form der Rationalität ist, wird Kapitel IV noch eingehen.)

Für die Handlungsplanung und das Selbstkonzept im HD sowie die Handlungsplanung im VD sind zwar noch statistisch bedeutsame, in der geforderten Richtung liegende, aber sicher eine sinnvolle Quantifizierung der Normwerte unterschreitende Wirkungen zu berichten. Die gegenüber dem VD im HD geringfügigeren Wirkungen auf das Selbstkonzept können auf die einseitigere Notenverteilung zurückgeführt werden, doch ist gleichzeitig ein deutlicherer Einfluß auf die Beliebtheit vorhanden. Das legt eine Interpretation mit Bezug auf die Ergebnisse der Voraussetzungsanalyse nahe: Dort wurde das Selbstkonzept im VD als eher auf die Prüfung zentriert, das des HD als eher auf Kompetenzerfahrungen aus dem Studium bezogen charakterisiert. Dieser Unterschied könnte eine Verarbeitung der Noten im VD über das Selbstkonzept bewirken, während die Noten im HD eher über die situationsgebundene emotionale Wertung des Fachs verarbeitet werden (in anderer Terminologie: Sie werden weniger der eigenen Fähigkeit zugeschrieben). Der im VD und HD unterschiedliche Einfluß auf das Selbstkonzept eigener Fähigkeit entspricht den Ergebnissen der geschilderten Untersuchung von FRETZ & ENGLE (1973), die ähnliche Einflüsse auf Studienanfänger beschränkt fand. Trotz dieser plausiblen Erklärungen ist auch nicht auszuschließen, daß die Selbstselektionen im HD für die Unterschiede verantwortlich sind, wenngleich ein inhaltlicher Zusammenhang

nicht einfach herleitbar scheint.

Den Normwerten wird allein darin klar entsprochen, daß die Einflüsse auf die Motivation deutlich unter denen auf das Selbstkonzept bleiben; das gilt aber nur für das VD.

Wirkungen der Gesamtnoten auf die Gesamtkonzepte

Wegen der kaum von der Notenstufe "1" abweichenden Gesamtnoten im HD kann die Auswertung nur unter Hinzunahme der Daten der Voruntersuchungen in der "Weiten Stichprobe" vorgenommen werden.

Tabelle: III.II. - 3

Wirkung der Gesamtnoten auf die Gesamtkonzepte (HD: "Weite Stichprobe")

	Vordiplom		HD (Weite Stichprobe)	
	PVF/PFF SK.GESAMT	PVF/PFF SK.ZUKUNFT	PVF/PFF SK.GESAMT	PVF/PFF SK.ZUKUNFT
part. Korrelation Gesamtnote/PFF (PVF auspart.)	+.31	+.34	+.06	+.40
Wahrscheinlichkeit zufälliger Abweich. von $r_{part.} = .00$	$<.05$	$<.05$	n.s.	$<.05$
Varianzanalyse mit Auspart. des PVF Wahrscheinlichkeit zufälliger Abweich.	n.s.	n.s.	n.s.	$(<.10)$
parameterfreier Vierfelder-Test Wahrscheinlichkeit zufälliger Abweich.	n.s.	n.s.	n.s.	n.s.

Methodisches Vorgehen: Wie in den beiden vorhergehenden Tabellen werden verschiedene Auswertungsverfahren parallel dargestellt. In den ersten beiden Zeilen erfolgt die Auswertung durch die Berechnung einer partiellen Korrelation. Dem schließt sich die Varianz-

analyse mit Auspartialisierung und die parameterfreie Auswertung mit dem Fisher-Yates-Test an. Positive Korrelationen entsprechen Wirkungen in der Richtung der Normwerte: erhöhtes Selbstkonzept bei besseren Noten. Wegen einzelner fehlender Antworten weicht die Zahl der in die Berechnung einbezogenen Prüflinge etwas von der Ausgangszahl VD: 43; HD ("Weite Stichprobe"): 46 ab.

Die Wirkung der Gesamtnoten auf die beiden Gesamtkonzepte im VD entspricht in der Höhe der partiellen Korrelation der Wirkung in den Einzelfächern. Sie kann statistisch weder in der Varianzanalyse noch in der parameterfreien Auswertung abgesichert werden.

Im HD wird ein Unterschied zwischen dem gegenwärtigen und zukunftsbezogenen Gesamtkonzept deutlich, der dem in den Normwerten postulierten Unterschied widerspricht. Wurde dort eine Einschränkung der Wirkung auf das gegenwärtige Selbstkonzept gefordert, beschränkt sich die empirische Wirkung auf das zukunftsbezogene Gesamtkonzept. Das Ausbleiben einer Wirkung auf das gegenwärtige Gesamtkonzept kann wie bei den Einzelfächern auf die aus Kompetenzerfahrungen im Studium zurückgehende Stabilisierung des Gesamtkonzepts zurückgeführt werden. Daß trotz des hohen Zusammenhangs beider Gesamtkonzepte im HD (s. Kapitel III.I.E.3.) das zukunftsbezogene Gesamtkonzept von den Noten beeinflußt wird, wird dadurch möglich, daß das zukunftsbezogene Gesamtkonzept situationsbedingt nicht nur vergangene Kompetenzerfahrungen, sondern auch konkrete Erwartungen integrieren muß. Die kaum vorhandene prognostische Validität der Noten aber läßt einen Einfluß der Noten auf solche Erwartungen nicht sinnvoll erscheinen.

Die Einordnung aller Ergebnisse dieses Teils, die vom Thema der Untersuchung her gesehen einen zentralen Stellenwert haben, erfordert die Auseinander-

setzung mit zwei Einwänden:

Erstens könnte die größere Wirkung der Noten auf das Selbstkonzept im VD auf bewußte Konformität mit normativem Einfluß (TEDESCHI 1972) zurückgehen. Die Prüflinge könnten versuchen, bestimmten Erwartungen durch eine bewußte Veränderung der Selbstkonzepte im PFF, für die ein Zusammenhang mit den Noten ihnen unmittelbar einsehbar ist, zu entsprechen. Aber es ist unwahrscheinlich, daß sich die Prüflinge an ihre Antworten im PVF erinnern, wenn sie den PFF ausfüllen. Außerdem wäre eine bewußte Verfälschung der Antwort eher in reaktiver Form zu erwarten. Es ist zu vermuten, daß, wenn überhaupt, Prüflinge, die, wie Einstellungsskalen zeigen (TIMMERBRINK 1971), eher negativ gegen Prüfungen eingestellt sind, zu verhindern suchen, daß den Prüfungen Wirkungen solcher Art nachgewiesen werden können.

Ein zweiter Einwand setzt an der ursächlichen Bedeutung der Noten an. Es ist durch das Auspartialisieren des Zusammenhangs mit den Antworten im PVF keineswegs gesichert, daß die Noten die tatsächliche Ursache der Wirkungen sind. Die Ergebnisse des theoretischen Teils, in die sich die Resultate der empirischen Untersuchung einpassen, tragen allerdings zur Absicherung bei, daß mit den Noten die wirklichen Ursachen benannt sind.

Zusammenfassung (III.II.B.1.)

Die globale Wirkung der Noten kann in den einzelnen "Fach-Prüfer-Kombinationen" wegen der Eigenart der Notenverteilungen nur im VD untersucht werden. Verschiedene Auswertungsarten zeigen, daß beim Selbstkonzept die Wirkung der Noten durchgehend in der erwarteten Richtung erfolgt und in einzelnen Kombinationen Signifikanz erreicht. Die Wirkung auf Handlungsplanung und Beliebtheit ist demgegenüber nur geringfügig. Das

Ergebnis bestätigt sich bei der Auswertung "Über-alle-Fächer": Im VD haben die Noten einen im Ausmaß geringen, aber deutlichen Einfluß im Sinne der Normwerte auf das Selbstkonzept, während der Einfluß auf die beiden Motivationsfragen zwar in der Richtung der Normwerte erfolgt, aber unbedeutend bleibt. Werden nur die psychologischen Fächer in die Auswertung einbezogen, ist kein Einfluß auf die Motivation mehr festzustellen. Im HD dagegen ist die Wirkung sowohl auf das Selbstkonzept als auch auf die beiden Motivationsfragen unbedeutend.

Die Transformation der Noten auf gleiche Streuungen und Mittelwerte in den einzelnen "Fach-Prüfer-Kombinationen" hat nur geringe Bedeutung für das Ausmaß der Wirkung.

Den Normwerten wird durch diese Ergebnisse voll nur darin entsprochen, daß im VD die Wirkung auf das Selbstkonzept größer ist als die auf die Motivation. Alle Wirkungen bleiben unter einem sinnvollen quantitativen Ausmaß. Als mögliche Ursache der im HD im Gegensatz zum VD geringeren Wirkung der Noten auf das Selbstkonzept kann dessen Verankerung im durch das Studium bedingten Erleben eigener Kompetenz gelten.

Die Wirkungen der Gesamtnoten auf die Gesamtkonzepte sind im VD ähnlich hoch wie die bei den Einzelnoten festgestellten Wirkungen. Im HD kann die Auswertung nur an der "Weiten Stichprobe" durchgeführt werden: Dort ist eine Wirkung auf das gegenwärtige Selbstkonzept nicht vorhanden, was wiederum auf dessen Kompetenzverankerung zurückführbar ist.

Die Ursächlichkeit der Noten für die Wirkungen ist durch die Auswertungsverfahren allein nicht zu sichern. Die Einpassung der Ergebnisse in Erkenntnisse des theoretischen Teils aber macht die Ursächlichkeit wahrscheinlicher.

III.II.B.2.: Geschlecht der Prüflinge und gewählte Prüfungsordnung (APO/NPO) als Rahmenbedingungen

Die Ergebnisse des Teils 1 schließen eine weitere Einschränkung nicht aus: Die Wirkungen können von Rahmenbedingungen abhängig sein. Die theoretischen Berichte legen die Mitwirkung solcher Rahmenbedingungen als situative oder personale Bedingungen nahe. Die Normwerte gehen von einer universellen bedingungsneutralen Wirkung der Noten aus. Eine empirische Kontrolle dieser Annahme ist deshalb angebracht. Sollten dagegen von den Rahmenbedingungen abhängige Wirkungen auftreten, stellt das gleichzeitig die Möglichkeit dar, den Prozeß der Beeinflussung näher zu bestimmen.

Als situative Rahmenbedingung wird die durch den Prüfling gewählte Prüfungsordnung einbezogen. Die Prüfungsordnung stellt in erster Linie eine unterschiedliche Umgebungsbedingung her. Da allerdings die Wahl der Prüfungsordnung im Untersuchungszeitraum den Prüflingen überlassen bleibt, ist nicht auszuschließen, daß personale Bedingungen an dieser Entscheidung beteiligt sind und eine mögliche Abhängigkeit vermittelnd bewirken. Als zweite Rahmenbedingung wird das Geschlecht der Prüflinge in die Auswertung einbezogen. Sowohl beim Selbstkonzept eigener Fähigkeit als auch bei der Motivation finden sich genügend Hinweise auf geschlechtsspezifische Prozesse. In bezug auf Leistungen ist das Geschlecht eine wesentliche Bedingung unterschiedlicher familiärer Sozialisation (HECKHAUSEN 1972) und wohl auch schulischer Sozialisation.

Abhängigkeit der Wirkung von der Wahl der Prüfungsordnung

Wesentliche Unterschiede zwischen der im Untersuchungszeitraum eingeführten neuen Prüfungsordnung und der seit vielen Jahren gültigen alten Prüfungsordnung

sind zumindest für die Zeit der Übergangsbestimmungen nur im VD feststellbar (s. Kapitel III.I.A.). Dort wirkt sich die Verlagerung der schriftlichen Hausarbeit ins HD (NPO) so aus, daß die Rückmeldung über das Ergebnis der Arbeit, die in der APO vor den mündlichen Prüfungen (und entsprechend vor dem PVF) erfolgt, wegfällt. Ausgerichtet auf diese besondere Situation können differenzierte Erwartungen formuliert werden: Es ist zu erwarten, daß bei Wegfall dieser vorgängigen Stabilisierung des Selbstkonzepts die Wirkung der mündlichen Prüfungen zunehmen kann. Das gegenwärtige Gesamtkonzept sollte davon betroffen sein, während das zukunftsbezogene Gesamtkonzept und die auf die Prüfungsfächer gerichteten Selbstkonzepte weniger Einfluß unterliegen dürften. Auch die Fragen nach der Motivation dürften keinen Einfluß zeigen.

Tabelle: III.II. - 4

Wirkung unter beiden Prüfungsordnungen im VD (APO/NPO)

	Alte Prüfungsordnung			Neue Prüfungsordnung			Wahrscheinlichkeit zuf.Abw. d.beiden part. Korrel.
	Anz. der Prüflinge	part. Korr.	Wahrsch. zuf.Abw. von $r_{part.} = .00$	Anz. der Prüflinge	part. Korr.	Wahrsch. zuf.Abw. von $r_{part.} = .00$	
Gesamtkonzepte							
Ges.Noten/ PFF-SK.GESAMT (PVF auspart.)	23	.47	<.05	18	.25		n.s.
Ges.Noten/ PFF-SK.ZUKUNFT (PVF auspart.)	23	.36		18	.38		n.s.

Fortsetzung der Tabelle s.S. 247

Tabelle: III.II. - 4 (Fortsetzung)

Wirkung unter beiden Prüfungsordnungen im VD (APO/NPO)

	Alte Prüfungsordnung			Neue Prüfungsordnung			Wahrscheinlichkeit zuf.Abw. d.beiden part. Korrel.
	Anz. der Prüfungen	part. Korr.	Wahrsch. zuf.Abw. von $r_{part.} = .00$	Anz. der Prüfungen	part. Korr.	Wahrsch. zuf.Abw. von $r_{part.} = .00$	
Über-alle-Fächer							
Noten/ PFF-SK.FÄCHER (PVF auspart.)	um 150	.34	<.01(?)	um 110	.24	<.05(?)	n.s.(?)
Noten/ PFF-MOT.W.A. (PVF auspart.)	um 150	.15		um 110	.13		n.s.(?)
Noten/ PFF-MOT.BEL. (PVF auspart.)	um 150	.20	<.05(?)	um 110	.11		n.s.(?)

Methodisches Vorgehen: Für die beiden Prüfungsordnungen sind jeweils die Zahl der Prüflinge (oder der Prüfungen, die durch fehlende Antworten etwas schwankt), die partielle Korrelation und deren Signifikanz angegeben. Die partielle Korrelation zeigt bei positivem Vorzeichen eine Wirkung im Sinne der Noten an: bei besseren Noten erhöhtes Selbstkonzept und vergrößerte Motivation. Die Prüfung auf Unterschiedlichkeit beider partieller Korrelationen erfolgt nach dem üblichen Verfahren des Vergleichs zweier Korrelationskoeffizienten (GUILFORD 1973), wobei wegen der Verwendung partieller Korrelationen die Zahl der Freiheitsgrade um jeweils 4 gegenüber der Zahl der Personen verringert wird, weil ein zusätzlicher Freiheitsgrad durch die Berechnung der partiellen Korrelationen belegt ist. Wenn an dieser Stelle noch einmal auf die Problematik inferenzstatistischer Auswertung der Werte "Über-alle-Fächer" verwiesen wird, hat das den besonderen Anlaß,

daß durch die Mitwirkung von bestimmten Personencharakteristika an der Entscheidung für eine bestimmte Prüfungsordnung die Personencharakteristika dadurch verstärkt in die Berechnungen Eingang finden können, daß sie über mehrere Einzelantworten, die für jede Person verrechnet werden, sich auswirken können.

Insgesamt gesehen sind die partiellen Korrelationen in beiden Gruppen von Prüflingen nicht sehr unterschiedlich. Keiner der Unterschiede ist nur annähernd statistisch signifikant. Der Unterschied beim gegenwartsbezogenen Selbstkonzept weist in die den Erwartungen entgegengesetzte Richtung. Zusammen mit den gleichsinnigen (aber auch geringfügigen) Unterschieden bei den Fächerfragen legt er die Vermutung nahe, daß er über Persönlichkeitsbedingungen, die die Wahl der Prüfungsordnung begründen, vermittelt ist. (Weitere Aufklärung könnte nur eine umfangreichere Erhebung bringen).

Abhängigkeit der Wirkung vom Geschlecht der Prüflinge

Wegen der vollständigen Abhängigkeit der einzelfachbezogenen Antworten über eine Personeneigenschaft wie das Geschlecht der Prüflinge ist eine Auswertung der einzelfachbezogenen Antworten nicht sinnvoll. Die Auswertung der Gesamtkonzepte kann nur für das VD erfolgen, da die Gesamtnoten des HD zu wenig streuen.

Tabelle: III.II. - 5

Wirkung auf Prüflinge unterschiedlichen Geschlechts im VD

	Männliche Prüflinge			Weibliche Prüflinge			Wahrscheinlichkeit zuf.Abw. d.beiden part. Korrel.	Wahrscheinlichk. zuf. Interaktion Geschlecht x Noten Varianzanalyse (PVF auspart.)
	Anz. der Prüflinge	part. Korr.	Wahrsch. zuf.Abw. von $r_{part.} = .00$	Anz. der Prüflinge	part. Korr.	Wahrsch. zuf.Abw. von $r_{part.} = .00$		
Gesamtkonzepte								
Ges.Noten/ PFF-SK.GESAMT (PVF auspart.)	26	.38	n.s.	15	.33	n.s.	n.s.	n.s.
Ges.Noten/ PFF-SK.ZUKUNFT (PVF auspart.)	26	.42	$<.05$	15	.16	n.s.	n.s.	n.s.

Methodisches Vorgehen: Die Tabelle entspricht dem Prinzip nach der vorhergehenden. Als zusätzliches Verfahren zur Prüfung einer Mitwirkung des Geschlechts wird die Varianzanalyse verwendet, bei der die Noten in zwei Stufen eines Faktors dichotomisiert sind und das Geschlecht als ein zweiter Faktor eingeführt wird; die Antworten des PVF werden aus den Antworten des PFF auspartialisiert.

Die Unterschiede der Wirkung auf die Gesamtkonzepte sind in beiden statistischen Verfahren nicht annähernd signifikant. Damit ist ein bedeutendes Ausmaß an Verfälschung der Wirkung der Noten nicht festzustellen, wenngleich wie im vorigen Abschnitt anzumerken ist, daß für solche Überprüfungen die vorliegenden Datensätze nicht ausreichen (beim zukunftsbezogenen Gesamtkonzept ist deskriptiv ein deutlicher Unterschied vorhanden).

Zusammenfassung (III.II.B.2.)

Die Normwerte fordern eine universelle Wirkung der Noten auf Selbstkonzept und Motivation. Die Mitwirkung bestimmter Rahmenbedingungen bei dem Einfluß der Noten ist aber nicht ausgeschlossen. Als solche Rahmenbedingungen werden im VD die gewählte Prüfungsordnung (APO und NPO) und für die Gesamtkonzepte des VD das Geschlecht der Prüflinge in die Auswertung einbezogen. Beiden Rahmenbedingungen ist keine statistisch bedeutsame "Mitwirkung" nachzuweisen. Eine bedingungsspezifische Verfälschung der Wirkung der Noten ist aber nicht auszuschließen, weil sich die Datensätze für eine Überprüfung nur beschränkt eignen.

III.II.B.3.: "Fach-Prüfer-Kombinationen" des VD als Rahmenbedingungen

Im Teil 1 stellte sich heraus, daß in einzelnen "Fach-Prüfer-Kombinationen" des VD sehr unterschiedliche partielle Korrelationen zwischen Noten und Selbstkonzept eigener Fähigkeit bestehen. Eine Signifikanzprüfung über alle Korrelationen hinweg konnte aber keine statistische Absicherung der Unterschiede erbringen. Trotzdem soll der Versuch gemacht werden, mögliche Bedingungen dieser Unterschiede zu ermitteln. Es liegt nahe, die Gründe für die Unterschiede in Eigenschaften der "Fach-Prüfer-Kombinationen" zu suchen, die aus den Antworten der Prüflinge deutlich werden. Ein einfaches Modell solcher Bedingungen geht von der Erwartung aus, daß lineare Zusammenhänge zwischen den Eigenschaften der "Fach-Prüfer-Kombinationen" und der Wirkung auf das Selbstkonzept bestehen. Sie können über Korrelationen zwischen den mittleren Einschätzungen der "Fach-Prüfer-Kombinationen" mit den als neuen Meßwerten betrachteten partiellen Korrelationen aus den einzelnen Kombinationen bestimmt werden. Diese Auswertung wird in Exkurs II als Ebene "Mittelwerte der Fach-Prüfer-Kombinationen" näher erläutert.

Tabelle: III.II. - 6

Zusammenhänge zwischen Eigenschaften der "Fach-Prüfer-Kombinationen" und dem Ausmaß ihrer Wirkung auf das Selbstkonzept

Wirkungsmaß:	part. Korrelation		PHI-Korrelation	
	Produkt-Moment-Korr.	Rang-Korr.	Produkt-Moment-Korr.	Rang-Korr.
PVF - F 3. FESTLEGUNG	.79+	.68	-.09	.18
PVF - F 7. PRÜFUNGSANGST	.51	.46	.62	.75+
PFF - F11. STUDIENLEISTUNG	-.72+	-.69+	-.73+	-.71+
PFF - F16. ANPASSUNG	-.73+	-.70+	-.46	-.38
Noten	-.35	-.21	-.69+	-.63

Methodisches Vorgehen: Als Ausgangswerte werden die in den 9 "Fach-Prüfer-Kombinationen" aus Teil 1 für die Antworten errechneten Mittelwerte verwendet. Zusätzlich wird das arithmetische Mittel der in den Kombinationen erteilten Noten einbezogen. Diese Werte werden mit den partiellen Korrelationen, die nach ihrer z-Transformation als neue Meßwerte betrachtet werden, korreliert. Als Verfahren kommen eine Produkt-Moment-Korrelation und zum Vergleich eine Rangkorrelation zur Anwendung, weil die z-transformierten partiellen Korrelationen sehr extreme Werte aufweisen. Parallel zu den partiellen Korrelationen wird als zweites Maß der Wirkung auf das Selbstkonzept die aus der parameterfreien Auswertung in der Vierfeldertafel berechnete PHI-Korrelation zwischen den nach "1" und "größer als 1" dichotomisierten Noten und den dichotomisierten Differenzwerten zwischen dem Selbstkonzept des PVF und des PFF verwendet. Auch diese Korrelationen werden z-transformiert mit den Mittelwerten nach beiden Verfahren korreliert. Die zweiseitige Signifikanzgrenze der Rangkorrelation mußte nach einem von LIENERT (1973) vorgeschlagenen Verfahren bestimmt werden. Korrelationen, die mit $p < .05$ signifikant sind, sind mit einem "+" kenntlich gemacht. Es werden alle Fragen aufgeführt, bei denen mit einer der vier Methoden mindestens ein mit $p < .05$ gegen Zufall abgesicherter Zusammenhang nachweisbar ist. Die Bedeutung der Richtung der Korrelationen wird in der Interpretation erläutert.

Die Eignung dieses Ansatzes, die situative Bedeutung der "Fach-Prüfer-Kombinationen" für das Ausmaß der Wirkung der Noten auf das Selbstkonzept eigener Fähigkeit zu untersuchen, ist in zweifacher Weise begrenzt. Eine prinzipielle Möglichkeit der "Scheinrealisation" geht darauf zurück, daß die Prüflinge in bestimmten Fächern den Prüfer wählen können. Wenn die Beeinflußbarkeit eines Prüflings mit den Determinanten

seiner Wahl konfundiert ist, kann eine Beziehung zwischen Mittelwerten und Ausmaß des Einflusses schon allein dadurch bedingt sein. Eine weitere Einschränkung betrifft die Zufälligkeit der Bedingungen: Dadurch, daß nur wenige besondere Prüfer prüfen und diese Prüfer den Fächern und damit den "Fach-Prüfer-Kombinationen" keineswegs zufällig zugeordnet sind, ist eine Verallgemeinerung der Ergebnisse über diese konkrete Prüferauswahl hinweg äußerst fragwürdig. Was in der Auswertung als abstrakte Eigenschaft der Situation erscheint, kann nur ein Aspekt der ganz konkret unterschiedlichen Persönlichkeiten der einzelnen an den Prüfungen beteiligten Prüfer sein. Aus beiden Gründen sind die folgenden Interpretationen nur erste Hinweise auf mögliche Zusammenhänge, die mit Hilfe eines geeigneteren Untersuchungsansatzes einer gezielteren Aufklärung bedürfen.

Die Ergebnisse nehmen dem zweiten Einwand etwas an Schärfe, da sich zu den Einschätzungen, die im PVF oder PFF die persönliche Eigenart des Prüfers zum Inhalt haben, keine bedeutsamen Beziehungen ergeben. Bei der Auswertung der Tabelle: III.II. - 6 ist als erstes festzustellen, daß die partiellen Korrelationen und die PHI-Koeffizienten deutlich voneinander abweichende Ergebnisse erbringen. Die Korrelationen der beiden Maße der Wirkung sind entsprechend nicht signifikant; die Rangkorrelation beider beträgt .51, die Produkt-Moment-Korrelation sogar nur .27. Die beiden Verfahren der Rangkorrelation und Produkt-Moment-Korrelation stimmen gut miteinander überein.

Trotz der geringen Beziehung der beiden Maße der Wirkungen zeigen sie bei einer Variablen gleichgerichtete und fast gleich enge Zusammenhänge an: zwischen dem Ausmaß der Wirkung und der mittleren Antwort der Prüflinge auf die Frage, ob die erhaltene Note eher in positiver oder eher in negativer Richtung von ihren Studienleistungen abweicht. Die von einem Prüfer verge-

bene Note hat dann einen um so größeren Einfluß auf das Selbstkonzept der Prüflinge, je besser sie im subjektiven Vergleich der Prüflinge mit den eigenen Studienleistungen ausfällt. Dementsprechend sind die Korrelationen des Ausmaßes an Einfluß mit den Notenmittelwerten der "Fach-Prüfer-Kombinationen" negativ (und in einer der vier Auswertungen sogar signifikant). Das bedeutet, daß die Noten in den "Fach-Prüfer-Kombinationen" den größeren Einfluß haben, in denen sie die weniger differenzierte Verteilung aufweisen, informationsstatistisch gesehen also den geringeren Informationsgehalt haben. Das Ergebnis ist mit Erkenntnissen des theoretischen Teils (Kapitel III.II.A.1) in Verbindung zu bringen, die nur bei Erfolg eine normgerechte Wirkung der Rückmeldung nahelegen, mit dem Zusatz, daß es sich hier um einen unerwarteten Erfolg handelt.

Die übrigen Korrelationen, die nur jeweils bei einem der Maße der Wirkung nachzuweisen sind, zeigen in eine Richtung, die durch den beschriebenen Zusammenhang mit der Erfolgsgetöntheit der Prüfungssituation vorgezeichnet wird. Der Einfluß auf das Selbstkonzept eigener Fähigkeit steigt in den "Fach-Prüfer-Kombinationen", in denen die Prüflinge weitgehende Freiheit in der Wahl des eigenen Prüfungsstoffes haben (PVF - F 3. FESTLEGUNG). Das ist mit der nachträglich in die Dissonanztheorie eingeführten Bedingung des "commitment" zu erklären (BREHM & COHEN 1962); die Dissonanz zwischen Selbstkonzept eigener Fähigkeit und abweichender Note wird nur erlebt, wenn Freiheit in der Kontrolle eigenen Handelns (Selbstbestimmung) erlebbar ist. Ganz ähnlich kann der Zusammenhang mit der nach der Prüfung erlebten Anpassung (PFF - F 16. ANPASSUNG) gedeutet werden. Dort, wo die Leistungen relativ frei von äußeren Beschränkungen wie Festlegungen des Stoffes und Zwängen der Situation sind, erhält die Note

als formulierte Bewertung des Prüfers eine größere funktionale Bedeutung. Das Bild der einflußreichen "Fach-Prüfer-Kombination" vervollständigt noch der Zusammenhang mit der erlebten Prüfungsangst (PVF - F 7. PRÜFUNGSANGST). Der Einfluß einer "Fach-Prüfer-Kombination" ist dann größer, wenn sie bei den Prüflingen weniger Angst erzeugt.

Alle Befunde zusammengesehen können auch so integriert werden, daß "liberalere"Prüfer durch ihre Noten mehr Einfluß haben. Die auf die Person der Prüfer zugeschnittene Interpretation macht die schon zu Anfang des Teils ausgesprochene Warnung aktuell, die Zusammenhänge nicht zu abstrakt zu sehen und davon ausgehend zu verallgemeinern; sie können in den Eigenarten der wenigen beteiligten Prüfer völlig an die einmalige Situation der Organisation der Prüfungen und Zusammensetzung des Prüferkollegiums gebunden sein.

Zusammenfassung (III.II.B.3.)

Über die Zusammenhänge zwischen den Maßen für das Ausmaß der Wirkung auf das Selbstkonzept der Prüflinge in den einzelnen "Fach-Prüfer-Kombinationen" des VD mit den mittleren Einschätzungen der "Fach-Prüfer-Kombinationen" durch die Prüflinge und den mittleren Noten sollen Rahmenbedingungen der Wirkung auf das Selbstkonzept ermittelt werden. Es zeigt sich, daß die Wirkung der Noten auf das Selbstkonzept um so größer ausfällt, je mehr die Note die Prüflinge auf eine unerwartet gute Leistung aufmerksam macht, mit der sie selbst auf Grund ihrer bisherigen Studienleistungen nicht rechneten. Weitere nicht so gut abgesicherte Zusammenhänge sind: Die Wirkung ist um so größer, je besser die Noten ausfallen, je freier die Wahl des Stoffes ist, je weniger sich die Prüflinge dem Prüfer anpassen mußten und je weniger Prüfungsangst sie in der Vorbereitung hatten. Die Möglichkeit, das als Syndrom eines li-

beralen Prüfers zu deuten, macht auf die Gefahr aufmerksam, Eigenschaften konkreter Personen in einer einmaligen Situation als abstrakte Situationsunterschiede mißzuverstehen. Eine Verallgemeinerung der Ergebnisse auf andere Situationen ist deshalb problematisch.

III.II.B.4.: Einfluß der Prüfung auf Kontingenzen zwischen den thematischen Fragen

Die bisherigen Auswertungen gehen allesamt davon aus, daß der Einfluß der Prüfung auf die thematischen Fragen über die Noten vermittelt ist. Es sind jedoch Einflüsse der Prüfungen denkbar, die unabhängig von den Noten wirken, zumal die wirksame Varianz der Noten in bestimmten "Fach-Prüfer-Kombinationen" sehr gering ist. Eine Möglichkeit, solche Änderungen zu erfassen, besteht darin, die Kontingenzen vor den Prüfungen (PVF) mit denen nach den Prüfungen (PFF) zu vergleichen.

Bei dieser Auswertung entsteht ein methodisch-statistisches Problem. Es ist kein statistisches Prüfverfahren vorhanden, das die Unterschiedlichkeit von zwei Korrelationen prüft, die paarweise voneinander abhängig sind. Es liegen nur Verfahren vor, die zwei über eine dritte Variable voneinander abhängige Korrelationen auf Unterschiedlichkeit prüfen. Da die aufwendigere varianzanalytische Prüfung nicht dem beschränkten Stellenwert der Auswertung entspricht, wird nur eine deskriptiv gestützte Interpretation der als Produkt-Moment-Korrelationen berechneten Kontingenzen vorgenommen.

Bei den erhaltenen Unterschieden zwischen den beiden Befragungen ist nicht mit Sicherheit festzustellen, in welchem Ausmaß beobachtete Veränderungen den Einflüssen der Noten, die in Teil 1 ermittelt wurden, entsprechen. Es soll versucht werden, die bisherigen Ergebnisse über den Einfluß der Noten in

Erwartungswerte umzusetzen, die dann mit den empirischen Korrelationen konfrontiert werden. Dabei ist es jedoch nur möglich, als Erwartungswerte eine Richtung der Änderung anzugeben, während quantitative Voraussagen erst ein endgültiges Urteil darüber erlauben würden, ob eine von den Noten unabhängige Wirkung der Prüfung behauptet werden kann.

Änderung der Korrelationen zwischen den thematischen Fragen

Zwischen den beiden Gesamtwerten des Selbstkonzepts eigener Fähigkeit sind für das HD auf Grund der Notenwirkungen keine Änderungen der Korrelationen zu erwarten, da die Gesamtnoten dort kaum streuen. Im VD ist dagegen durchaus möglich, daß die gleichsinnige Wirkung der Noten auf das gegenwärtige und zukunftsbezogene Gesamtkonzept eine Zunahme der Korrelationen im PFF zur Folge hat. Für die einzelfachbezogenen Antworten ist im HD auf Grund der Notenwirkung nur mit geringfügigen Änderungen der Korrelationen zu rechnen, während im VD die bedeutsamere Wirkung der Noten auf das Selbstkonzept eine Verringerung der Korrelationen zwischen Selbstkonzept und Handlungsplanung bzw. Beliebtheit verursachen kann.

Tabelle: III.II. - 7

Änderung der Korrelationen zwischen den thematischen Fragen

Einzelwerte	PVF			PFF		
	Fach-Prüfer-Kombinationen mittl. r	von bis	Über-alle-Fächer	Fach-Prüfer-Kombinationen mittl. r	von bis	Über-alle-Fächer
Vordiplom						
SK.FÄCHER/ MOT.BEL.	-.34	-.12 -.70	-.42	-.18	.10 -.64	-.20
SK.FÄCHER/ MOT.W.A.	-.15	.13 -.55	-.20	-.12	.20 -.49	-.13
MOT.BEL./ MOT.W.A.	.52	.29 .74	.55	.56	.04 .79	.69

Tabelle: III.II. - 7 (Fortsetzung)

Änderung der Korrelationen zwischen den thematischen Fragen

Einzelwerte	PVF			PFF		
	Fach-Prüfer-Kombinationen mittl. r	von bis	Über-alle-Fächer	Fach-Prüfer-Kombinationen mittl. r	von bis	Über-alle-Fächer
Hauptdiplom						
SK.FÄCHER/ MOT.BEL.	-.41	-.01 -.65	-.43	-.21	.10 -.67	-.25
SK.FÄCHER/ MOT.W.A.	-.32	.29 -.93	-.45	-.18	.29 -.72	-.26
MOT.BEL./ MOT.W.A.	.53	.15 .76	.58	.75	.01 .96	.76

Gesamtkonzepte	PVF	PFF
Vordiplom		
SK.GESAMT/ SK.ZUKUNFT	.51	.59
Hauptdiplom		
SK.GESAMT/ SK.ZUKUNFT	.71	.76

Methodisches Vorgehen: Die Berechnungen erfolgen, wie zu Tabelle: III.I. - 21 beschrieben. Zusätzlich sind die Korrelationen der Gesamtkonzepte angegeben, die über die Prüflinge hinweg berechnet sind. Die Zahl der Prüfungen beträgt im HD 145, im VD schwankt sie wegen fehlender Antworten um 280. Die Zahl der Prüflinge bei den Gesamtkonzepten weicht wegen fehlender Antworten etwas von der Gesamtzahl (VD: 43; HD: 29) ab. Wegen der andersartigen Polung zeigen negative Korrelationen zwischen Selbstkonzept und Motivation einen Zusammenhang: größere Motivation - höheres Selbstkonzept an.

Bei den Gesamtkonzepten sind jeweils nur geringfügige Erhöhungen der Korrelationen festzustellen, was den Erwartungen grob gesehen noch entspricht. Ein zusätzliches Ergebnis ist der engere Zusammenhang der

beiden Gesamtkonzepte im HD, der die Prüfung überdauert. In Kapitel III.I.E. wird der schon im PVF bestehende Unterschied zum VD interpretiert.

Die geringfügigen Unterschiede der über die "Fach-Prüfer-Kombinationen" hinweg gemittelten Korrelationen und der "Über-alle-Fächer" berechneten Korrelationen bestätigen noch einmal, daß die Kontingenzen im Antwortverhalten im wesentlichen durch Unterschiede der Prüflinge und individuelle Prüfungssituationen determiniert sind, während Prüfer und Prüfungsfach nur wenig dazu beitragen. Dabei können Prüfer und Prüfungsfach, wie die Streubereiche der Korrelationen zeigen, durchaus als Rahmenbedingung die Kontingenzen mitbedingen; sie tragen aber wenig systematische Varianz zu den Kontingenzen zwischen den Antworten bei.

Vor der Auswertung der Korrelationen der einzelfachbezogenen Antworten sind einige Überlegungen über die normative Einordnung der Zusammenhänge der Antworten angebracht. Sie erstellen Normwerte für mögliche Änderungen. Bei den beiden Motivationsfragen ist ein möglichst enger Zusammenhang als Kongruenz von Handlung und Tätigkeit sinnvoll. Die kongruente Einbettung gegenwärtiger Beliebtheit in überdauernde Handlungsperspektiven ist für persönliche Verwirklichung und sachliche Leistung ausschlaggebend. Für die Motivationsfragen und das Selbstkonzept ist ebenfalls, in Ergänzung der in Kapitel II nur als Erwartung formulierten Aussage, eine Vergrößerung des im PVF nicht sehr engen Zusammenhangs zu fordern. Die Passung von Fähigkeiten und Anforderungen muß nicht allein als Abstimmung von individuellen Fähigkeiten und gesellschaftlichen Anforderungen, sondern auch intrapersonal als Abstimmung von Fähigkeit und Motivation verstanden werden. Sowohl der handlungsperspektivische Person/Umwelt-Bezug als auch der gegenwärtig-emotionale Bezug sollten ein bedeutsames Ausmaß an Zusammenhang mit dem Selbstkonzept eigener Fähigkeit aufweisen. Es ist

anzunehmen, daß solche Übereinstimmungen zur Stabilität der Persönlichkeit und ihres aktiven Umwelt-Bezuges beitragen.

Im Sinne dieser Normwerte sind die von den Wirkungen der Noten her nicht erwarteten größeren Übereinstimmungen von Handlungsplanung und Beliebtheit nach der Prüfung als ein positiver Beitrag der Prüfung zu sehen. Eine alternative Interpretation dagegen führt die Zunahme lediglich auf den Wegfall der Prüfung als "Störbedingung" zurück, die die sinnvolle Konvergenz zwischen der handlungsperspektivischen und der mehr emotionalen Zuwendung durch die Einführung der Prüfung als neues Umweltelement nur vorübergehend auflöst. Für diese Erklärung spricht, daß die Zunahme des Zusammenhangs beim VD auf dem Einfluß der "Fach-Prüfer-Kombinationen" beruht. Die Anlage der Befragung und die unzureichende statistische Kontrollierbarkeit der Auswertung geben diesen Interpretationen lediglich den Charakter von Hinweisen.

Die Zusammenhänge von Selbstkonzept und Motivation ändern sich im VD, wie es auf Grund der Notenwirkung zu erwarten ist. Da jedoch im HD ähnliche Verringerungen der Korrelationen über die Prüfung hinweg vorkommen, ist die Änderung nicht als Folge der Notenwirkung (rein statistisch besteht bei derartig geringen Korrelationen keine Notwendigkeit, Notenwirkung und Änderung der Korrelationen als verbunden anzusehen), sondern als Einfluß der Prüfung, der unabhängig von den Noten erfolgt, zu erklären. Mit Bezug auf die neu formulierten Normwerte für solche Änderungen wäre das als negativer Beitrag der Prüfung zu sehen, indem die Entkoppelung von Selbstkonzept und Motivation eine Labilisierung der Persönlichkeitsstruktur mit sich bringt. Auch bei dieser Bewertung ist wiederum eine Alternativerklärung möglich, die den Zusammenhang vor der Prüfung allein als "Einmischung" der Motivation in das Selbstkonzept begreift (s. Kapitel III.I.E.3).

Da im HD jedoch auch der Zusammenhang von Handlungsplanung und Selbstkonzept von dieser Destrukturierung betroffen wird, legt das unabhängigere kompetenzorientierte Selbstkonzept im HD die anfängliche Interpretation nahe. Auf jeden Fall aber ist festzuhalten, daß die Prüfung keinen Beitrag zur Stabilisierung der Person über eine Abstimmung von Selbstkonzept und Motivation leistet, wie die Normwerte es fordern.

Änderung der Korrelationen innerhalb der thematischen Fragen

Für die drei thematischen Fragen im HD sind keine bedeutenden Änderungen zu erwarten, da die Noten dort wenig Wirkungen aufweisen (s. Teil 1). Im VD ist nur im Selbstkonzept eigener Fähigkeit eine Änderung möglich. Doch auch sie dürfte geringfügig sein; zudem ist ihre Richtung nicht eindeutig anzugeben. Die Korrelationen zwischen den Noten in den einzelnen Fächern (s. Kapitel III.I.B.3.) sind nicht deutlich von den Korrelationen zwischen den Selbstkonzepten im PVF unterschieden. Demnach sollte eine Wirkung der Noten keine Änderung der Korrelationen zwischen den Selbstkonzepten nach sich ziehen. Die in den "Fach-Prüfer-Kombinationen" unterschiedliche Wirkung kann u.U. sogar zu einer Verringerung der Korrelationen führen.

Tabelle: III.II. - 8

Änderung von Streubereich und Mittelwerten der Korrelationen der einzelfachbezogenen Antworten

		SK.FÄCHER	MOT.BEL.	MOT.W.A.
Vordiplom				
PVF	mittl. Korr.	+.38	+.07	+.05
Streubereich		-.15 bis +.64	-.13 bis +.43	-.37 bis +.56
PFF	mittl. Korr.	+.51	+.16	-.02
Streubereich		+.36 bis +.71	-.09 bis +.49	-.44 bis +.55
Hauptdiplom				
PVF	mittl. Korr.	+.32	-.05	-.11
Streubereich		+.16 bis +.56	-.29 bis +.37	-.65 bis +.55
PFF	mittl. Korr.	+.56	+.17	-.02
Streubereich		+.37 bis +.74	-.07 bis +.64	-.26 bis +.31

Methodisches Vorgehen: Die Berechnungen entsprechen Tabelle: III.I. - 19.

Die beiden Motivationsfragen unterliegen keinen auffallenden Änderungen. Nur im HD entdifferenziert sich nach der Prüfung die Gegenstandsstruktur, die die Handlungsplanung begründet: Der Streubereich der Korrelationen zwischen den einzelnen Fächern wird geringer. Weitere Nachforschungen müßten bestätigen, ob das als Folge einer unsicheren Berufsperspektive gedeutet werden kann.

Das fachbezogene Selbstkonzept eigener Fähigkeit ändert sich in beiden Prüfungsarten gleichsinnig in einer Richtung, die durch die Erwartungen nicht vorgezeichnet wird: Die Zusammenhänge zwischen den Selbstkonzepten für die einzelnen Fächer sind nach der Prüfung enger als vor der Prüfung. Diese Entdifferenzie-

rung kann in einer persönlichkeitsspezifischen und damit für alle einzelnen Konzepte gleichen Verarbeitung der Erfahrung der Prüfung begründet sein, die unabhängig von den Noten erfolgt und eine individuelle Reaktion auf das "Bestehen" der Prüfung ist. Da die beiden Gesamtkonzepte schon einen engeren Zusammenhang aufweisen, tritt bei ihnen diese Entdifferenzierung nicht so deutlich hervor. Zu den persönlichkeitsspezifischen Bedingungen, die an einer derartigen Reaktion beteiligt sein können, gehört die individuelle Ausprägung des Selbstwertgefühls. Eine Entdifferenzierung der Struktur des Selbstkonzepts kann vonvorübergehender Dauer sein. Auf Grund der Befragung sind darüber keine Erkenntnisse vorhanden. Selbst dann kann sie Entscheidungen, die in der Zeit nach der Prüfung über die weitere Ausbildung oder Tätigkeit zu treffen sind, erschweren, wenn jene z.B. die Wahl eines spezifischen Arbeitsgebietes zum Inhalt haben. Dieser Einfluß der Prüfung ist deshalb negativ zu bewerten.

Zusammenfassung (III.II.B.4.)

An Änderungen der Korrelationen zwischen den thematischen Fragen, die vom PVF zum PFF hin erfolgen, können Wirkungen der Prüfungen abgelesen werden, die nicht über die Noten vermittelt sein müssen. Die bisher festgestellten Wirkungen der Noten lassen kaum beträchtliche Änderungen der Korrelationen zwischen den thematischen Fragen erwarten. Als Normwerte für solche Änderungen werden gefordert: Beliebtheit und Handlungsplanung sollten einen engen Zusammenhang aufweisen. Auch Selbstkonzept und Motivation sollten aufeinander abgestimmt sein, weil das zu personaler Stabilität beiträgt und einen aktiven Umwelt-Bezug ermöglicht. Der Zusammenhang zwischen Beliebtheit und Handlungsplanung nimmt über die Prüfung hinweg zu. Es ist aber möglich, daß das lediglich der Wegfall der Prüfung

als "Störbedingung" bewirkt. Der Zusammenhang zwischen Motivation und Selbstkonzept nimmt ab. Das kann nicht allein auf die Notenwirkung zurückgeführt werden und ist auf jeden Fall den Normwerten entgegengesetzt.

Auch Änderungen der Korrelationen zwischen den Einzelfächern innerhalb der einzelnen thematischen Fragen können solche Wirkungen außerhalb der Noten anzeigen. Die Korrelationen innerhalb Handlungsplanung und Beliebtheit ändern sich wenig. Die Korrelationen zwischen den fachbezogenen Selbstkonzepten nehmen deutlich zu. Das kann in der persönlichkeitsspezifischen Verarbeitung der Prüfungserfahrung begründet sein und ist als Entdifferenzierung unrealistisch, Entscheidungen eher hinderlich und im normativen Rahmen negativ zu bewerten.

III.II.B.5.: Einfluß der Prüfung auf die Antwortverteilungen der thematischen Fragen

Ein weiterer Schritt der Auswertung, der wie der vorangehende Teil solche Einflüsse der Prüfung nachzuweisen versucht, die nicht in den Kontingenzen mit den Noten erfaßt werden, knüpft an das Eingangszitat an. Es sind Einflüsse der Prüfung denkbar, die sich auf alle Prüflinge gleich, zumindest auf alle Prüflinge in der gleichen Richtung auswirken. Solche Einflüsse müßten sich in Änderungen der Mittelwerte oder Streuungen der Urteilsverteilungen zwischen den beiden Befragungen niederschlagen.

Auf der Grundlage der Einwirkung der Noten allein sind keine Änderungen zu erwarten, wenngleich einige Erkenntnisse aus der theoretischen Analyse von Änderungsprozessen für gerichtete Verschiebungen der Mittelwerte sprechen: z.B. die Asymmetrie der Reaktion auf Erfolg und Mißerfolg beim Selbstkonzept für eine weitere Zunahme des Mittelwerts der Antworten. Demgegenüber müßte ein Beitrag der Rückmeldung zur Realistik der

Selbstkonzepte eine Annäherung der Mittelwerte an die objektive Skalenmitte bewirken, wenn der Anspruch aufrechterhalten wird, daß die Antworten darin zumindest verteilungsorientiert erfolgen sollten. Für die Motivationsfragen ist die theoretische und empirische Basis für Formulierung von Erwartungen und Normwerten nicht ausreichend.

Tabelle: III.II. - 9

Änderung der Mittelwerte und Varianzen über die Prüfung hinweg

	Zahl der Prüflinge/ Prüfungen	Mittelwerte PVF	Mittelwerte PFF	Varianzen PVF	Varianzen PFF	Wahrscheinlichk. zuf. Unterschiede Mittelwerte	Wahrscheinlichk. zuf. Unterschiede Varianzen
Vordiplom							
Gesamtkonzepte							
SK.GESAMT	41	5.48	5.88	2.01	3.49	<.05	<.05
SK.ZUKUNFT	41	6.10	6.33	2.43	2.34	n.s.	n.s.
Werte Über-alle-Fächer							
SK.FÄCHER	273	5.32	5.96	4.00	4.48	<.01(?)	
MOT.W.A.	283	2.90	2.84	.89	.96		
MOT.BEL.	275	2.40	2.51	.92	1.05	<.05(?)	
Hauptdiplom							
Gesamtkonzepte							
SK.GESAMT	27	6.33	6.43	3.15	3.74	n.s.	n.s.
SK.ZUKUNFT	27	5.74	6.11	3.35	1.72	n.s.	<.01
Werte Über-alle-Fächer							
SK.FÄCHER	145	5.54	6.25	4.39	3.99		
MOT.W.A.	145	2.37	2.32	1.07	1.01		
MOT.BEL.	145	2.20	2.24	.88	1.05		

Methodisches Vorgehen: Die Mittelwerte und Varianzen werden mit t-Tests für abhängige Stichproben (GUILFORD 1973) auf Abweichungen geprüft. Die Signifi-

kanzangaben für die "Über-alle-Fächer" berechneten Parameter sind wegen der partiellen Abhängigkeit nur bedingt gültig. Dortfehlt deshalb auch die Angabe "n.s.", da sie gleichfalls nur diese bedingte Gültigkeit beanspruchen kann. Höhere Mittelwerte bedeuten ein höheres Selbstkonzept und mindere Motivation.

Die Gesamtkonzepte eigener Leistungsfähigkeit nehmen in beiden Prüfungsarten übereinstimmend über die Prüfung hinweg zu. Nur beim gegenwärtigen Selbstkonzept im VD ist diese Zunahme signifikant. Ohne Vergleich mit Ergebnissen unter anderen Bedingungen ist nicht sicherzustellen, ob die Mittelwertverschiebung auf die besondere Eigenart der Notenverteilungen zurückgeht oder eine Wirkung des "Bestehens" der Prüfung ist. Selbst als letzteres ist sie nicht eindeutig, da sie sowohl eine zusätzliche Überschätzung als auch nur der Ausgleich einer durch die Erwartung der Prüfung bedingten perspektivischen Unterschätzung vor der Prüfung sein könnte. Dessenungeachtet leistet die Prüfung im Hinblick auf die Normwerte keinen Beitrag zur Realistik der Selbstbewertung.

Die gleichzeitig bei den Gesamtkonzepten vorhandenen Varianzunterschiede sind in ihrer Bedeutung ebensoschwer einzuschätzen, u.a. weil sie im VD und HD in unterschiedliche Richtungen weisen. Im HD ist eine Verringerung der Varianz entsprechend der kaum vorhandenen Streuung der Gesamtnoten verständlich; allerdings ist sie hier auf das zukunftsbezogene Selbstkonzept beschränkt. Einen "positiven" Beitrag der Prüfung aber kann man in dieser Veränderung der Varianz kaum sehen.

Die Parameteränderungen in den Selbstkonzepten der Einzelfächer bringen keine neuen Aspekte. Bei Handlungsplanung und Beliebtheit sind die Verschiebungen der Parameter insgesamt gesehen geringfügig. Das einzige statistisch (mit Bedenken) abgesicherte Ergebnis besagt, daß im VD nach der Prüfung die Beliebtheit der

Fächer im Mittel abnimmt. Eine dissonanztheoretische Interpretation des Ergebnisses könnte vermuten: Die Beliebtheit der Fächer vor der Prüfung wird aufgewertet, weil die Prüflinge sich mit einem bestimmten Grad an Freiwilligkeit der Prüfung als einer an sich unangenehmen Handlung unterworfen haben und die dadurch entstehende Dissonanz durch eine Aufwertung der Fächer verringern können. Nach der Prüfung fällt diese dynamische Bedingung fort, und die Beliebtheit der Fächer sinkt entsprechend. Eine andere Interpretation scheint weniger spekulativ zu sein. Da die Frage im PVF auf das Studium gerichtet ist, wird dort die immanente Anreizbedingung der Fachinhalte angesprochen, während die Frage im PFF die gegenwärtige Beliebtheit erfragt, die nach der Prüfung der Zukunftsperspektive angeglichen wird. Die Handlungsplanung ändert sich im Mittel nicht. Eine Auslegung der Theorie intrinsischer Motivation, die von einer Verringerung in Situationen äußerer Belohnung ausgeht (s. Kapitel III.II.A.2.) ist zumindest an den kognitiven Antizipationen solcher intrinsischen Motivation nicht zu belegen.

Die Veränderungen der Mittelwerte und Varianzen in den einzelnen "Fach-Prüfer-Kombinationen" bestätigen zwar im großen und ganzen die Ergebnisse der Gesamtwerte, doch zeigen sie außerdem noch Unterschiede zwischen den einzelnen Kombinationen auf, die aber im Rahmen dieser Untersuchung so wenig abgeklärt werden können, daß eine Darstellung nicht erfolgt.

Zusammenfassung (III.II.B.5.)

Von den Noten unabhängige Wirkungen der Prüfungen lassen sich auch an der Verschiebung der Parameter der Antwortverteilungen über die Prüfung hinweg aufweisen. Die Selbstkonzepte nehmen im Mittel über die Prüfung hinweg zu. Das ist ein weiterer Schritt in Richtung fehlender Realistik, dessen Ursache aber nicht eindeutig

zu interpretieren ist. Änderungen der Varianz der Antworten zum Selbstkonzept, bei den Mittelwerten der Beliebtheit und weitere unterschiedliche Verschiebungen in den einzelnen "Fach-Prüfer-Kombinationen" deuten komplexere Wirkungsprozesse an, die im Rahmen dieser Untersuchung nicht aufzuklären sind.

III.II.B.6.: Strukturelle Einbettung der thematischen Fragen

Die beiden Fragebögen erfassen viele unterschiedliche Aspekte des Erlebens und Verhaltens der Prüflinge. Die Einbettung der thematischen Fragen in das Gesamt der übrigen Fragen trägt zur Aufklärung der Hintergründe der bisherigen Ergebnisse bei.

Anwendung der Faktorenanalyse als Methode der Strukturanalyse

Die Faktorenanalyse stellt eine Suchstrategie zur Aufdeckung komplexerer Beziehungen in einem Variablengesamt dar, das so umfangreich ist, daß allein Beziehungen zwischen einzelnen Variablen eine mögliche Ordnungsstruktur nicht mehr erkennbar machen. Dabei ist die immer wieder übersehene Selbstverständlichkeit zu betonen, daß die erhaltene Struktur nicht die Struktur des Gegenstands darstellt, sondern von der Vorgabe der Variablen bestimmt wird. Insofern ist zu fordern, daß der Anwendung der Faktorenanalyse die entsprechende Absicht bei der Datenerhebung vorauszugehen hat. Das ist in dieser Untersuchung nur bedingt der Fall; der Erkundungscharakter der Untersuchung kommt dieser Forderung mit der Formulierung vieler aspektverschiedener einzelner Fragen jedoch entgegen. Die Aspektvielfalt ist besonders bei den einzelfachbezogenen Fragen vorhanden. Die Anlage als prae-postdesign mit der Wiederholung bestimmter Fragen, u.a. der thematischen, gibt ein mögliches Grundmuster fak-

torieller Strukturierung vor, in dem die thematischen Fragen mit den übrigen vergleichbar sind.

Die Entscheidung, welche Ebene der Kovarianzen als Grundlage der Faktorisierung dienen soll, hat von den methodisch begrenzten Möglichkeiten eines Vergleichs von unterschiedlichen Faktorenanalysen auszugehen. Als zugleich grundlegende Ebene der bisherigen Auswertung bieten sich die "Über-alle-Fächer" berechneten Korrelationen an, die jede einzelne mündliche Prüfung als Datenträger betrachten. Die bisherigen Auswertungen zeigen, daß die Kovarianzen zwischen den Antworten im wesentlichen durch die personale, an die konkrete Situation der einzelnen Prüfung gebundene Varianz determiniert werden. Sie werden nicht vorwiegend durch personale Konstanten bedingt. Es ist selbstverständlich wie in den bisherigen Auswertungen eine getrennte Analyse für die beiden Prüfungsarten (VD und HD) notwendig.

Das methodische Vorgehen sieht die Berechnung von Produkt-Moment-Korrelationen zwischen den thematischen Fragen (SK.FÄCHER, MOT.W.A., MOT.BEL.) und den übrigen "Fächerfragen" (s. Anhang I) vor. Auch die Noten werden miteinbezogen. Die Extraktion der Faktoren wird nach der Hauptkomponentenmethode vorgenommen. Die erhaltenen Faktoren werden dann mit einer Anzahl von 2 Faktoren beginnend bis zu der Zahl, die durch das Abbruchkriterium: Eigenwert >1 festgelegt wird, varimax-rotiert (ÜBERLA 1968).

Die Darstellung der rechnerischen Ergebnisse bezieht die Ergebnisse aller durchgeführten Rotationen ein. Der Vergleich der Rotationen mit unterschiedlichen Faktorenzahlen erlaubt ein Urteil über die Stabilität von einzelnen Faktoren. Auf die Wiedergabe der zahlenmäßigen Größe der Ladungen wird verzichtet; zur besseren Überschaubarkeit sind die Ladungen unterschiedlicher Größe zu zwei Gruppen zusammengefaßt, die durch entspre-

chende Symbole repräsentiert werden:

Ladungen von .50 bis .69: o

Ladungen ab .70: O

Die Vorzeichen der Ladungen sind in den Symbolen enthalten. Sie ergeben sich aus der Verschlüsselung der Fragen, die von links beginnend mit 1, 2, 3 usw. die Skalenstufen in Zahlen umsetzt (s. Anhang I).

Faktorenstruktur im VD

In der Hauptkomponentenextraktion weisen 5 Faktoren einen Eigenwert >1 auf; sie erklären zusammen 53.5 % der Gesamtvarianz. Auf die einzelnen unrotierten Faktoren entfallen 22.8, 11.6, 8.4, 5.8, 4.9 % der Varianz (der Eigenwert >1 entspricht 3.8 % der Varianz).

Tabelle: III.II. - 10 (s.S. 271)

Rotierte Faktorenstruktur der Korrelationen "Über-alle-Fächer" im VD

Methodisches Vorgehen: Die Reihenfolge der Fragen entspricht nicht der im Fragebogen. Die thematischen Fragen sind vorgezogen. Die Darstellung berücksichtigt die Ladungen bei Rotationen mit 2, 3, 4 und 5 Faktoren.

Der Faktor I umfaßt in allen Rotationen den größten Varianzanteil. Inhaltlich gesehen bleibt er über die Rotationen hinweg im wesentlichen stabil. In diesem Faktor weisen solche Fragen eine bedeutsame Ladung auf, die den wertenden Person/Umwelt-Bezug der Prüflinge, der sich auf die Sachstruktur der Fächer richtet, erfassen. Darunter fallen sowohl die thematischen motivationalen Fragen als auch die Beurteilungen der Wichtigkeit der einzelnen Fächer. Die Benennung des Faktors könnte als fachbezogene Motivation erfolgen. Dieser vom Studium her schon stabilisierte

Tabelle: III.II. - 10

Rotierte Faktorenstruktur der Korrelationen "Über-alle-Fächer" im VD

	Faktor I				Faktor II				Faktor III				Faktor IV				Faktor V			
	Rotationen				Rotationen				Rotationen				Rotationen				Rotationen			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
PVF-SK.FÄCHER										⊖	●	●								
PFF-SK.FÄCHER																				●
PVF-MOT.W.A.	⊖	⊖	⊖	⊖																
PFF-MOT.W.A.	⊖	⊖	⊖	⊖																
PVF-MOT.BEL.	⊖	●	●	●																
PFF-MOT.BEL.	⊖	⊖	⊖	⊖																
NOTEN					⊖	⊖	●	●							●					
PVF-F 1.ATMOSPHÄRE										●	●	●								
PVF-F 2.ANFORDERUNGEN												●								
PVF-F 3.FESTLEGUNG																				
PVF-F 4.ANNEHMLICHKEIT	●									●	●	●								
PVF-F 5.RELEVANZ	⊖	⊖	⊖	⊖																
PVF-F 6.WICHTIGKEIT	⊖	⊖	⊖	⊖																
PVF-F 7.PRÜFUNGSANGST										⊖	⊖	●								
PVF-F 8.NOTENERWARTUNG										⊖	⊖	●								
PFF-F 9.WICHTIGKEIT	●	⊖	⊖	⊖																
PFF-F 10.WIRKL.LEISTUNG					●	●	⊖	⊖												
PFF-F 11.STUDIENLEISTUNG					●	⊖	⊖	⊖												
PFF-F 12.RÜCKM.NEG.																				●
PFF-F 13.RÜCKM.POS.															●	●				
PFF-F 14.KONGRUENZ.VORB.															●					●
PFF-F 15.KONGRUENZ.STUD.															●	●				
PFF-F 16.ANPASSUNG																				
PFF-F 17.ATMOSPHÄRE					●	●									⊖	⊖				
PFF-F 18.ANFORDERUNGEN																				
PFF-F 19.PRÜF.INSGESAMT					●	●									⊖	⊖				
Anteile an der Gesamtvarianz (%)	22.0	18.7	18.3	18.2	12.4	11.5	8.9	9.2	-	12.6	12.7	12.3	-	-	8.8	8.2	-	-	-	5.7

Person/Umwelt-Bezug wird durch den Vollzug der Prüfung kaum tangiert. Er ist auch von der emotionalen Wertung der Vorbereitung unabhängig, wenn man von einer Ladung der Annehmlichkeit bei der 1. Rotation absieht. Rein formal sind Prüfung samt Vorbereitung und wertender Fachbezug nicht kongruent. Das Fehlen einer Ladung im Selbstkonzept zeigt die geringfügige Abstimmung von erlebter Fähigkeit und rationaler Handlungsplanung. Doch ist das eher der Instabilität des Selbstkonzepts anzulasten, während der wertende Fachbezug durch die Integration des emotionalen Bezugs (Beliebtheit) sich als stabil auszeichnet.

Der Faktor II erfährt im Laufe der Rotationen eine Verschiebung seiner Bedeutung, die eine Integration der Darstellung mit der des Faktors IV erfordert, zu dem ein Teil der Ladungen abwandert. Zusammen mit dem Faktor IV deutet er an, daß das Erleben der Prüfung nach deren Vollzug insgesamt in zwei verschiedene Aspekte zerfällt. Über die Rotationen hinweg zunehmend stellt der Faktor II den Aspekt der Gerechtigkeit der Leistungsbewertung heraus, die als Ergebnis (Produkt) der Prüfung sich vom Verlauf der konkreten Interaktion (Prozeß) ablöst. Die Verarbeitung des Ergebnisses der Prüfung als in der Prüfung erteilte Note erfolgt im wesentlichen in einer Stellungnahme zu dieser Note. Unter Verwendung des Begriffes der Zuschreibung (Attribution) spricht dieser Faktor die prüferattribuierte Verarbeitung durch den Prüfling an (Gerechtigkeit der Note). Eine Selbstzuschreibung als Einfluß auf das Selbstkonzept fehlt.

Die Verarbeitung der Leistungsbewertung in der Prüfung ist, wie die ersten beiden Rotationen zeigen, nicht unabhängig von der wertenden Bewältigung des Verlaufs der Prüfung selbst. Dieses prozeßorientierte Erleben der Prüfung als Gelingen der Interaktion, der Faktor IV, beinhaltet bei der nachträglichen Be-

wertung der Prüfungsinteraktion durch den Prüfling eine Mischung von einem rationalen didaktischen Bezugssystem, wie beispielsweise der positiven Rückmeldung in der Prüfung oder der Kongruenz der Prüfung, mit einem mehr emotionalen Bezugssystem (Annehmlichkeit der Prüfung). In einer der Rotationen und bei der anfänglichen Integration in den Faktor II weist der Faktor einen Zusammenhang mit der erteilten Note auf, indem bei besseren Noten das jeweils positivere Nacherleben der Prüfung stattfindet. Dafür bietet sich neben einer Interpretation, die die Noten als ursächlich für das Werten der Prüfungsinteraktion ansieht, auch eine weitere Interpretation an, die die Ursächlichkeit umkehrt, indem das Gelingen der Prüfungsinteraktion als Grundlage der Note betrachtet wird (SCHLIEHE & SEEGER 1974). Das wäre ein Hinweis auf eine mögliche Rationalität des Prüfungsvorgangs, die allerdings darin ihre Grenze findet, daß sie weder mit dem stabilen Person/Umwelt-Bezug noch mit dem leistungsbezogenen Selbsterleben der Prüflinge eine Verbindung findet. Die Prüfung ist auch dann, wenn sie unter partiell sachlichem Anspruch erlebt wird (Faktor IV) von studiumsbezogenen (fachbezogenen) personalen Strukturen relativ losgelöst.

Das Erleben der Prüfung zerfällt ohne Ausnahme in einen Aspekt, der die Zeit vor der Prüfung charakterisiert, und einen, der die Zeit nach der Prüfung charakterisiert, ohne daß zwischen beiden wie beim fachbezogenen Person/Umwelt-Bezug eine Kontinuität festzustellen wäre. Das zeigt über eine Dissoziierung von Prüfung und Studium hinaus eine emotional erlebnismäßige Dissoziierung des Prüflings-Prüfungs-Bezugs vor und nach der Prüfung. Das zukunftsperspektivische Erleben der Prüfung, der Faktor III, das im Gegensatz zum nachherigen Erleben Leistungsaspekte und Aspekte der Interaktion ungetrennt enthält, bleibt über drei

Rotationen hinweg stabil. In Absetzung vom eher sachlich getragenen motivationalen Person/Umwelt-Bezug umfaßt es solche Fragen, die einen situativ auf die Prüfung ausgerichteten emotionalen Bezug beinhalten, der von den sachlichen Inhalten abgelöst ist. Es fällt dabei auf, daß in diesen auf die isolierte Prüfungssituation abgestimmten Bezug auch das Selbstkonzept eigener Fähigkeit vor der Prüfung integriert ist, dem sich die Erwartung über die leistungsbezogene Bewältigung dieser Situation (Anspruchsniveau) zugesellt. Ein Zusammenhang mit der fremdbewerteten Leistung in der Prüfung (Noten) fehlt.

Der Faktor V ist vom Anteil aufgeklärter Varianz her gesehen nicht sehr bedeutend; er gewinnt an Bedeutung aber dadurch, daß in ihm als einzigem das Selbstkonzept eigener Fähigkeit nach der Prüfung eine Ladung aufweist. Das Selbstkonzept eigener Fähigkeit ist eingebettet in eine Struktur, die als selbstbewußtes Erleben der Bewältigung der Prüfung bezeichnet werden kann. Das Erleben erfolgreicher Bewältigung der Prüfung geht einher mit positivem Einfluß auf das Selbstkonzept. Dieses Erleben, das aus Fehlen negativer Rückmeldung und der Kongruenz von Vorbereitung und Prüfungsinhalten besteht, zeigt aber keine Beziehung zu der Leistungsbewertung in der Prüfung. Zudem ist das Erleben der Bewältigung auf die Vermeidung von Mißerfolg eingegrenzt und bemißt den sachlichen Gehalt der Prüfung nur an der Vorbereitung auf die Prüfung, nicht aber am Studium selbst.

Faktorenstruktur im HD

In der Hauptkomponentenextraktion weisen wie beim VD 5 Faktoren einen Eigenwert >1 auf; sie erklären zusammen 54.5 % der Gesamtvarianz. Auf die einzelnen unrotierten Faktoren entfallen 23.1, 11.2, 9.2, 6.6 und 4.3 % der Varianz (der Eigenwert >1 entspricht 3.8 %

der Varianz).

Die Darstellung versucht, sich an die Interpretation der Ergebnisse für das VD anzulehnen. Das wird dadurch erleichtert, daß in groben Zügen eine Übereinstimmung der beiden Analysen festzustellen ist, die als Bestätigung dafür angesehen werden kann, daß die verwendeten Variablen bedeutsame Strukturbeziehungen aufweisen und als solche eine valide Interpretationsgrundlage sind. Die Identität geht wiederum nicht so weit, daß die gleichen Benennungen der Faktoren zutreffend sind. Auf eine Neuformulierung der Benennungen aber wird verzichtet.

Tabelle: III.II. - 11 (s.S. 276)

Rotierte Faktorenstruktur der Korrelationen "Über-alle-Fächer" im HD

Methodisches Vorgehen: s. Tabelle: III.II. - 10.

Der Faktor I des motivierten Fachbezuges stellt auch hier den wichtigsten Faktor dar. Es kommen aber zusätzlich zu den Ladungen im VD noch einige Ladungen hinzu, die die Bedeutung dieses Faktors etwas modifizieren. Auch die erlebte Annehmlichkeit der Vorbereitung lädt jetzt in allen Rotationen. Bedeutsamer ist, daß das Selbstkonzept eigener Fähigkeit des PVF in den Faktor integriert wird. In der Vorbereitung auf das HD ist es dem fachbezogenen Person/Umwelt-Bezug möglich, zu den Inhalten der Vorbereitung ein ähnlich sachliches Verhältnis zu finden wie zu den Inhalten des Studiums selbst. Die Integration des Selbstkonzepts erfolgt auf der Grundlage dessen Verankerung im Erleben sachlich begründeter Kompetenz (s. oben). Die sachliche Stabilisierung des Person/Umwelt-Bezuges erweist sich sogar tendenziell (in den ersten beiden Rotationen) als Bezugssystem zur Einschätzung der in der Prüfung erfolgten Leistungsbewertung tragfähig. Im

Tabelle: III.II. - 11

Rotierte Faktorenstruktur der Korrelationen "Über-alle-Fächer" im HD

	Faktor I				Faktor II				Faktor III				Faktor IV				Faktor V			
	Rotationen				Rotationen				Rotationen				Rotationen				Rotationen			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
PVF-SK.FÄCHER	●	●	●	●																●
PFF-SK.FÄCHER																				●
PVF-MOT.W.A.	⊖	⊖	⊖	⊖																
PFF-MOT.W.A.	⊖	⊖	⊖	⊖																
PVF-MOT.BEL.	●	●	●	⊖																
PFF-MOT.BEL.	⊖	⊖	⊖	⊖																
NOTEN					●	●	●													●
PVF-F 1.ATMOSPHÄRE					●				●	●	●									
PVF-F 2.ANFORDERUNGEN									●	●										
PVF-F 3.FESTLEGUNG																				
PVF-F 4.ANNEHMLICHKEIT	●	●	●	●																
PVF-F 5.RELEVANZ	⊖	⊖	⊖	⊖																
PVF-F 6.WICHTIGKEIT	⊖	⊖	⊖	⊖																
PVF-F 7.PRÜFUNGSANGST									⊖	⊖	⊖									
PVF-F 8.NOTENERWARTUNG									●	●										
PFF-F 9.WICHTIGKEIT	⊖	⊖	⊖	⊖																
PFF-F10.WIRKL.LEISTUNG							●	●												
PFF-F11.STUDIENLEISTUNG	●	●					●	●												
PFF-F12.RÜCKM.NEG.																				
PFF-F13.RÜCKM.POS.																				
PFF-F14.KONGRUENZ.VORB.																●				
PFF-F15.KONGRUENZ.STUD.															⊖	●				
PFF-F16.ANPASSUNG																				
PFF-F17.ATMOSPHÄRE					⊖	●	●	●				●								
PFF-F18.ANFORDERUNGEN															⊖	●				
PFF-F19.PRÜF.INSGESAMT					⊖	⊖	⊖	⊖												
Anteile an der Gesamtvarianz (%)	23.0	22.7	21.6	20.4	11.4	10.8	9.8	9.6	-	10.0	10.2	9.7	-	-	8.6	6.3	-	-	-	8.6

Laufe der Rotationen allerdings verselbständigt sich dieser Aspekt ähnlich wie beim VD.

Der Faktor II gibt die individuell-situative Verarbeitung der Leistungsbeurteilung in der Prüfung wieder. Während im VD der Faktor sich im Laufe der Rotationen in die beiden Aspekte der Verarbeitung der Leistungsbewertung und der Bewältigung der Prüfungsinteraktion aufspaltet, findet im HD eine Vereinigung beider Aspekte zu einem Faktor statt, der mit der Leistungsbewertung sowohl eine Beurteilung der Angemessenheit der Bewertung als auch der Annehmlichkeit der Situation verbindet.

Der Faktor III stellt das an der konkreten Situation ausgerichtete Erleben der Prüfung im PVF dar; allerdings lädt das Selbstkonzept eigener Fähigkeit nicht mehr in diesem Faktor. Der Faktor konzentriert sich ganz auf die Antizipation der Prüfungssituation; er verliert seine Bedeutung für die zuständlich emotionale Wertung der Prüfungsvorbereitung, die im Gegensatz zum VD als sachlicher Bezug bewertet wird.

Der Faktor IV ist nicht mit dem entsprechenden Faktor des VD zu vergleichen. Er stellt vielmehr ein Urteil über den inhaltlichen Verlauf der Prüfung dar, bei dem eine höhere Kongruenz der Prüfungsinhalte mit höheren Anforderungen einhergeht. Ein sachliches Bezugssystem löst damit das mehr emotionale, auf das Gelingen der Prüfungsinteraktion gerichtete des VD ab.

Der vom aufgeklärten Varianzanteil her gesehen nicht mehr so bedeutende Faktor V ist ebenfalls im VD nicht vorhanden. Er hat insofern eine große inhaltliche Bedeutung, als das Selbstkonzept eigener Fähigkeit in ihm sowohl vor als auch nach der Prüfung eine bedeutsame Ladung aufweist. Ein Teil der Varianz des Selbstkonzepts besteht aus einer im Hinblick auf die Prüfung realistischen Einschätzung der eigenen Fähigkeit, die deshalb auch die Prüfung überdauern kann.

Bei der Einordnung der Ergebnisse der Faktorenanalysen muß die Gefahr im Auge behalten werden, die Filterung von Strukturen aus der Beziehung zwischen einzelnen Variablen als Nachweis von Kausalitäten mißzuverstehen. Insgesamt gesehen aber passen die Ergebnisse der Faktorenanalysen zu den Ergebnissen bei der Analyse der Wirkungen. Die Stabilität der fachlich orientierten Person/Umwelt-Bezüge über die Prüfung hinweg entspricht der geringen Wirkung der Prüfung auf diese Bezüge, zumal dann, wenn wie beim VD die Prüfung als ein zweifach isoliertes Ereignis erscheint: als vom Studium losgelöst und getrennt in ein Erleben der kommenden Prüfung und eine Verarbeitung der Prüfungserfahrung. Im HD gilt nur das zweite Merkmal der Isolation, während dort das Erleben der Vorbereitung auf die Prüfung teilweise in die fachorientierten Bezüge integriert wird.

Der geringen Wirkung der Noten auf das Selbstkonzept eigener Leistungsfähigkeit entspricht im VD über die Isolierung der Prüfung hinausgehend die Verarbeitung der Noten in einer sachlichen und einer mehr emotionalen, auf die Prüfungsinteraktion bezogenen, Bewertung der Prüfung durch den Prüfling. Im HD steht einer Wirkung neben der sachlichen Bewertung der Prüfung (als einer Alternative zur Selbstzuschreibung) eine gewisse realistische Kontinuität des Selbstkonzepts über die Prüfung hinweg entgegen, was noch einmal die bisherigen Hinweise auf die Kompetenzorientierung des Selbstkonzepts im HD bestätigt.

Zusammenfassung (III.II.B.6.)

Eine Faktorenanalyse der Korrelationen zwischen den fachbezogenen Fragen, die auf der Ebene "Über-alle-Fächer" berechnet werden, soll für das VD und das HD getrennt Informationen über die Einbettung der thematischen Fragen in das Gesamt der Fragen liefern. Die

Faktoren werden nach der Hauptkomponentenmethode extrahiert und einer Varimax-Rotation unterzogen. Nach dem Eigenwertkriterium sind in beiden Analysen 5 Faktoren bedeutsam. Sie erfassen im VD 53.5 % und im HD 54.5 % der Gesamtvarianz.

Eine Interpretation der Faktoren, die auch ihre Stabilität über die Rotationen hinweg einbezieht, identifiziert im VD einen zentralen Faktor fachbezogener Motivation, der über die Prüfung hinweg konstant bleibt. Zwei weitere Faktoren gelten der Verarbeitung der Erfahrung der Prüfung: der Gerechtigkeit der Leistungsbewertung und dem Gelingen der Prüfungsinteraktion. Von dem nachträglichen Erleben der Prüfung losgelöst ist ein Faktor des emotional zentrierten zukunftsperspektivischen Erlebens der kommenden Prüfung, in dem auch das Selbstkonzept vor der Prüfung eine Ladung aufweist. Ein weiterer Faktor, dem das Selbstkonzept nach der Prüfung zugeordnet ist, kann als selbstbewußtes Erleben der Bewältigung der Prüfung gesehen werden, ist aber von den erteilten Noten unabhängig.

Die Faktorenstruktur des HD zeigt grob gesehen Übereinstimmung, weicht aber in Einzelheiten deutlich von der des VD ab. Wiederum vorhanden ist der Faktor fachbezogener Motivation, dem sich neben dem sachlichen Erleben der kommenden Prüfung auch das Selbstkonzept zugesellt. Ein zweiter Faktor faßt die Verarbeitung der Leistungsbewertung mit der durch den Prüfling vorgenommenen Bewertung des Prüfungsverlaufs insgesamt zusammen. Neu ist ein Faktor, der die Bewertung der Prüfung in einem sachlichen Bezugssystem zum Inhalt hat. Wiederum ist das Erleben der kommenden Prüfung von dem Erleben der bewältigten Prüfung fast vollständig geschieden. Im HD ist das Erleben der kommenden Prüfung mehr als im VD rein auf die emotionalen Aspekte beschränkt. Ein fünfter Faktor deutet eine gewisse Realistik des Selbstkonzepts und dementsprechend mögliche

Kontinuität über die Prüfung hinweg an.

Die bisherigen Ergebnisse der Wirkungsanalyse passen sich gut in die Ergebnisse der Faktorenanalyse ein.

III.III.: Stufenanalyse

Übersicht: Das Kapitel beginnt mit einer begrifflichen und methodischen Festlegung des Vorgehens der Stufenanalyse. In der Stufenanalyse werden Prozesse isoliert, die vermittelnd oder alternativ in die globale Wirkung der Noten auf Selbstkonzept und Motivation eingreifen können. Dabei liegt der Schwerpunkt der Analyse im empirischen Vorgehen.

Die Struktur der subjektiven Notenskalen kann in die Wirkung der Noten auf Selbstkonzept und Motivation vermittelnd eingreifen. Zugleich ist sie selbst möglicherweise abhängig von den in der Prüfung erteilten Noten. Teil 1 beschäftigt sich mit den über Nutzen-Skalierung gewonnenen subjektiven Abständen zwischen den einzelnen Noten.

Eine weitere Stufe des Wirkungsprozesses ist dem Kontext der Leistungsmotivationsforschung entnommen: das Anspruchsniveau. Es dient hier der Bestimmung des Ausmaßes subjektiven Erfolgs oder Mißerfolgs in der Prüfung. In Teil 2 soll geprüft werden, ob der subjektive Erfolg oder Mißerfolg im Vergleich zum objektiven eine bedeutsamere Wirkung auf Selbstkonzept und Motivation hat.

Ebenfalls aus dem Rahmen der Psychologie leistungsmotivierten Handelns stammt der Begriff der erlebten Selbstverantwortlichkeit für das Ergebnis der Prüfung. In Teil 3 wird die Bedeutung der Selbstverantwortlichkeit für die Wirkung der Noten untersucht.

Der 4. Teil gilt der erlebten Rückmeldung in den einzelnen mündlichen Prüfungen, bei der die posi-

tiv zentrierte Rückmeldung von der negativ zentrierten unterschieden wird. Beide können vermittelnd oder alternativ der Wirkung der Noten sich zugesellen.

Inhaltliche und methodische Struktur der Stufenanalyse

Die Stufenanalyse geht davon aus, daß die Umsetzung der Fremdbewertung in Selbstkonzept und Motivation über ein komplexes intrapersonales Geschehen verwirklicht wird, in dem objektive Information subjektiv verarbeitet wird. Subjektive Notenskala, Anspruchsniveau, Selbstverantwortlichkeit und erlebte Rückmeldung sind Bestandteile eines solchen Prozesses. (Auch soziale Prozesse sind an der Verarbeitung beteiligt; sie sind nicht Gegenstand der Untersuchung.)

Es soll nicht versucht werden, wie der Begriff der Stufe es vermuten läßt, daraus ein Ablaufmodell zu konstruieren. Das wäre eine Überforderung der die einzelnen Stufen begründenden Theorien und Begriffe wie zugleich der Planung der Untersuchung, den Daten und den bisherigen Ergebnissen nicht angemessen. Die punktuelle Befragung reicht zur empirischen Verifizierung eines Ablaufmodells nicht aus. Die Stufen als Antworten im Fragebogen sind nur Aspekte eines komplexen, interaktiven Gesamtprozesses, in dem einfache Ursache-Wirkung-Zusammenhänge kaum klar heraustrennbar sind und deshalb hypothetisch bleiben (ein genereller Nachteil derartiger korrelativer Felduntersuchungen gegenüber experimenteller Variablenkontrolle). Um die hypothetischen Grundlagen nicht durch weitere Vorannahmen zu überlasten, werden deshalb die Stufen jeweils getrennt analysiert.

Ihre Stellung in bezug auf die Wirkungsanalyse ist in drei Formen möglicher "Wirkung" zu klassifizieren: 1. Die Stufen können alternative abhängige Variablen, also Objekte der Wirkung darstellen. Das gilt für die subjektive Notenskala, die wie Selbstkonzept und

Motivation sich in Abhängigkeit von den Noten ändern kann. 2. Die Stufen können alternativ zu den Noten als unabhängige Variablen gesehen werden, die Änderungen in Selbstkonzept und Motivation bewirken. Dafür kommen das Anspruchsniveau und die erlebte Rückmeldung in Frage. 3. Es ist auch möglich, daß die Stufen als Moderatorvariablen in die Wirkung der Noten auf die thematischen Variablen eingreifen. Das gilt für die subjektive Notenskala, die Selbstverantwortlichkeit und die erlebte Rückmeldung.

Wenn bei der zweiten und dritten Form der Verwendung eine Wirkung angenommen wird, ist das eher final als kausal zu verstehen. Zusammenhänge von Stufen und Änderungen der thematischen Variablen oder die Interaktion von Noten und Stufen bei der Wirkung auf die thematischen Variablen qualifizieren einen Gesamtprozeß, bei dem der Aspekt der Änderung von Selbstkonzept und Motivation "finalisiert" wird. Im Rahmen dieser Arbeit ist es ausreichend, Bedingungen nachzuweisen, die einen von den Noten unabhängigen oder deren Wirkung modifizierenden Zusammenhang mit Selbstkonzept und Motivation deutlich werden lassen. Die Frage nach der Kausalität ist demgegenüber zweitrangig, solange nicht aus den Ergebnissen praktische Schlußfolgerungen für gezielte Eingriffe gezogen werden. Die bedürften der Begründung in kausalen Analysen.

Mit Ausnahme der erlebten Rückmeldung weisen die Stufen eine Gemeinsamkeit auf: Sie stammen aus dem weit abgesteckten Bereich der Psychologie leistungsmotivierten Verhaltens. Als Bestandteile motivationaler Prozesse sind sie Anzeichen für Wirkungen von Prüfungen (bzw. Noten), die nicht im Sinne der Normwerte erfolgen. (Es kann insgesamt gesehen kaum positiv gewertet werden, daß sich die Motivationspsychologie unter Verwendung dieser "Stufen" im wesentlichen mit den Bedingungen "atypischen" Verhaltens beschäftigt und die Un-

tersuchung von Bedingungen, unter denen typisches Verhalten erfolgt, vernachlässigt. Für den Stellenwert der Stufenanalyse ist allerdings dieser Ansatz geeignet.)

Für das methodische Vorgehen ist eine Festlegung der Auswertungsebenen der fachbezogenen Fragen notwendig. Da sich in der Wirkungsanalyse herausgestellt hat, daß die Auswertung "Über-alle-Fächer" grob gesehen die Ergebnisse der einzelnen "Fach-Prüfer-Kombinationen" zutreffend zusammenfaßt, soll die Auswertung der einzelfachbezogenen Daten im wesentlichen auf dieser Ebene erfolgen. Die inferenzstatistischen Probleme bleiben dabei bestehen: Es werden partiell abhängige Daten verwendet.

Als Auswertungstechniken kommen im wesentlichen die Berechnung partieller Korrelationen und Kovarianzanalysen zur Anwendung; auch hier gelten dieselben Einschränkungen wie für die Wirkungsanalyse: Die Voraussetzungen der Methoden sind nicht voll erfüllt.

III.III.1.: Subjektive Notenskalen der Prüflinge

Die Verwendung subjektiver Notenskalen geht davon aus, daß die Abstände zwischen den einzelnen Notenstufen bei den einzelnen Prüflingen unterschiedlich sein können (s. Kapitel III.I.B.1.) und daß diese Unterschiede als Moderatorvariable für die Wirkung der Noten in Frage kommen. Mögliche Änderungen der subjektiven Notenskalen in Abhängigkeit von den erhaltenen Noten könnten Alternativen für die Verarbeitung der Noten in Änderungen der thematischen Variablen darstellen.

Subjektive Notenskalen können wegen der aufwendigen Befragungsmethode (PVF-II/PFF-III.NOTENSKALA) nur einmal und nicht für jede mündliche Prüfung differenziert ermittelt werden und finden deshalb nur im Zu-

sammenhang mit den Gesamtnoten und den Gesamtkonzepten Berücksichtigung. Die Auswertung bezieht auch solche Aspekte ein, die nicht der Verwendung als Moderatorvariable oder als alternative abhängige Variable zuzuordnen sind.

Theorie, Methode und funktionale Bedeutung subjektiver Notenskalen

Ausgangspunkt des Verfahrens zur Bestimmung subjektiver Notenskalen ist der aus der Spieltheorie stammende Ansatz der Nutzenskalierung (SEU-Skalierung). Das Verfahren wird in methodischen Vorarbeiten von SIEGEL und anderen aus der Spieltheorie abgeleitet und auf die Analyse des Anspruchsniveaus bei Noten übertragen (BECKER & SIEGEL 1958; SIEGEL 1957). Es werden jeweils zwei Paare von Noten vorgegeben, zwischen denen der Urteilende auszuwählen hat. Aus den Wahlen kann eine Skala subjektiver Nützlichkeiten und entsprechend der Abstände zwischen den Noten entwickelt werden. Dafür eignen sich zwei verschiedene Verfahren. Bei BECKER & SIEGEL wird eine gleiche Wahrscheinlichkeit der Noten von .50 durch die Instruktion suggeriert. In diesem Fall geben die Präferenzen für eines der beiden Notenpaare direkt die relative Nützlichkeit wieder. Am Beispiel des ersten Notenpaares (s. Anhang I) "1/3 - 2/2" bedeutet das: Für den, der die Alternative "1/3" wählt, ist die "1" gegenüber der "2" nützlicher als die "2" gegenüber der "3". In eine Skala umgesetzt, die die Nützlichkeit in Form von Abständen zwischen den Noten ausdrückt, läßt die Wahl der Alternative "1/3" eine subjektive Notenskala ableiten, bei der die "1" von der "2" weiter entfernt ist als die "2" von der "3". Werden systematisch verschiedene Paarungen konstruiert und die Präferenzen nach diesem Verfahren geordnet, lassen sich die Abstände zwischen den Noten in eine Ordnungsrelation bringen (ordered metric scale) und

auch Kombinationen der Abstände ordnen (higher ordered metric scale).

Zur Kritik bietet das Verfahren an zwei verschiedenen Stellen Anlaß: Zum einen kann bezweifelt werden, daß durch einfache Instruktion die Wahrscheinlichkeiten für die Noten, die in der von den Urteilenden erlebten Wirklichkeit ganz andere Werte haben, erlebnismäßig so beeinflußt werden können, daß die Urteilsgrundlage eine "erlebte" Gleichwahrscheinlichkeit ist. Es ist anzunehmen, daß die empirischen Wahrscheinlichkeiten von Noten, mit denen die Urteilenden in ihrer gegenwärtigen Situation konfrontiert sind, in den Entscheidungsprozeß eingehen (HERBST 1972). Zum anderen stellen die einzelnen Spiele gegen den Zufall in den meisten Fällen solche Paare zusammen, bei denen eine der Paarungen mit Sicherheit eine Note verspricht. Damit aber wird subjektiv eine Wahl verlangt zwischen Risiko und Sicherheit. Es ist nicht auszuschließen, daß dadurch Persönlichkeitseigenschaften wie eine allgemeine Risikobereitschaft angesprochen werden und das spezifische Risiko überlagern.

Ein von diesem abweichendes Verfahren wird von McDANIEL et al. (1961) vorgeschlagen, indem eine sichere Note gegen die Wahl eines Notenpaares aus einer schlechteren und einer besseren Note gesetzt wird, für das die komplementären Wahrscheinlichkeiten systematisch variiert werden. Aus der Wahrscheinlichkeit, bei der die Wahl von der sicheren Note zu dem ungleichen Notenpaar umschlägt, kann auf Grund der Annahme einer Nützlichkeitsfunktion unter Verwendung der Notenwahrscheinlichkeiten der Abstand der Noten berechnet werden. In ähnlicher Weise gehen auch SCHMIDT et al.(1971) vor, nachdem SCHMIDT (1966) in einer vorausgehenden Arbeit die erstere Methode angewendet hat. Auch bei diesem Verfahren sind wie bei dem ersteren die beiden kritischen Fragen zu stellen.

Für die Verwendung im Rahmen dieser Arbeit ist zu bestimmen, welche psychischen Prozesse durch die Methode zur Ermittlung subjektiver Notenskalen angesprochen werden. Eine nicht ganz befriedigende Antwort geben einige Versuche zur Validierung des Verfahrens.

Bei BECKER & SIEGEL (1958) wird die subjektive Notenskala in engem Zusammenhang mit dem Anspruchsniveau (zum Begriff s. auch Teil 2) gesehen; es wird bestimmt als die höhere Note der beiden, die innerhalb der subjektiven Skala den größten Abstand voneinander aufweisen. BECKER & SIEGEL berichten eine hohe Übereinstimmung von r = .84 zwischen dem aus Interviewdaten gewonnenen Anspruchsniveau und dem aus der Skala ermittelten Anspruchsniveau. Dieser Nachweis der Validität wird durch eine weitere Untersuchung beider Autoren (BECKER & SIEGEL 1962) bestätigt, in der durch manipulierte Notenerteilung Erfolg oder Mißerfolg erzeugt wird. Die subjektiven Notenskalen zeigen Veränderungen, die dem manipulierten Erfolg und Mißerfolg gleichlaufen: Nach Erfolg werden die über Notenskalen ermittelten Anspruchsniveaus erhöht, nach Mißerfolg findet eine Senkung statt. Auch die Ergebnisse von SCHMIDT et al. (1971) mit dem zweiten Verfahren der Vorgabe unterschiedlicher Wahrscheinlichkeiten sprechen dafür, daß das individuelle Anspruchsniveau an der Konstitution der Skalen beteiligt ist: Die Beliebtheit von Fächern weist eine Beziehung zu dem für sie ermittelten Notennutzen auf; gleichzeitig korrelieren die wirkliche Leistung und , unter bestimmten Bedingungen, der fremdeingeschätzte Ehrgeiz von Schülern mit deren Notenskalen. Die situative Bedingtheit der Notenskalen bei SCHMIDT (1966) kann über Änderungen des Anspruchsniveaus vermittelt sein. In einer Situation, in der die subjektiven Notenskalen für "unbedeutende" Noten ermittelt werden, werden individuell sehr unterschiedliche Skalen verwirklicht. Wird dagegen der Situation durch die In-

struktion Ernstcharakter dadurch verliehen, daß die Noten als Prüfungsergebnis zählen sollen, entstehen übereinstimmend Skalen, deren größter Abstand zwischen der Notenstufe "4" und "5" liegt.

Dieses Ergebnis aber läßt auch Zweifel daran aufkommen, daß das konkrete Leistungsziel der Personen sich damit auf die Notenstufe "4" konzentriert. Weitere Ergebnisse anderer Untersuchungen entsprechen gleichfalls nicht der Interpretation, daß die subjektive Notenskala auf das Anspruchsniveau zurückzuführen sei. Der geringfügige Zusammenhang, der nach McDANIEL et al. (1961) zwischen den Notenskalen und der erreichten Abschlußnote besteht, wäre für das Anspruchsniveau höher zu erwarten. STALLINGS et al. (1969), die die subjektiven Notenskalen zur Entscheidung für ein differenziertes gegenüber einem zweistufigen Notensystem in Beziehung setzen, finden nur unbedeutende nichtsignifikante Korrelationen. Auch HERBST (1972) kann einen bedeutsamen Zusammenhang zwischen erwarteten oder angezielten Noten und den subjektiven Notenskalen bei Schülern nicht feststellen.

Diese Unklarheit ist Anlaß, eine Zurechnung zum Prozeß der Anspruchsniveausetzung nicht vorzunehmen, sondern unschärfer den die subjektiven Skalen begründenden Prozeß als Erfolgs-/Mißerfolgsorientierung zu bestimmen. Die Erfolgs-/Mißerfolgsorientierung wird als Rechts- oder Linksverschiebung gegenüber der gleichabständigen Skala deutlich (diese im folgenden weiterhin verwendete Beschreibung bezeichnet Skalen, die linksverschoben zunehmend größere Abstände zwischen den schlechteren Noten und rechtsverschoben zunehmend größere Abstände zwischen den besseren Noten zeigen). In einer RASCH-Skalierung weist HERBST (1972) nach, daß diese Dimension die wesentliche Gemeinsamkeit von 11 der 15 Spiele (Paarungen) ist. Eine Verbindung der Erfolgs-/Mißerfolgsorientierung zur entsprechenden Un-

terscheidung in der Leistungsmotivationsforschung ist nur grob begrifflich herstellbar und soll deshalb unterbleiben.

Ein praktisch verwendbares Maß für die Erfolgs-/Mißerfolgsorientierung, die als quantitativ abgestufte Dimension verstanden wird, wird modellfrei gewonnen, indem die Antworten zu jedem Spiel mit 1 und 2 verschlüsselt werden. Die Zuordnung der Zahlenwerte zu einem der beiden Paare ist objektiv entscheidbar: Das Paar mit den extremeren Noten liegt jeweils in der Richtung einer - abstrakt verstandenen - Rechtsverschiebung (Erfolgsorientierung). Eine Berechnung von Nutzenwerten, wie sie von SCHMIDT (1966) und überflüssigerweise auch von SCHMIDT et al. (1971) vorgenommen wird, ist nicht notwendig. Diese modellfreie Verwendung hat den zusätzlichen Vorteil, daß dabei die fragwürdigen Annahmen über die erlebte Notenwahrscheinlichkeit nicht impliziert sind. Bei der Berechnung der Gesamtwerte, die eine Analogie zur LIKERT-Skalierung der Einstellungsmessung darstellt, entsteht ein Problem durch die Möglichkeit einer unentschiedenen Antwort. Sie wird mit 1.5 verrechnet.

Da die Ergebnisse der RASCH-Skalierung den Anspruch erheben, von der konkreten Stichprobe unabhängig zu sein, werden die Ergebnisse von HERBST (1972) als Grundlage der Summierung über die 11 einer Dimension zugehörigen Spiele genommen. Die Spiele gehen aus der folgenden Anordnung hervor, die der Anordnung der Spiele in der Frage (s. Anhang I) entspricht; für die in die Rechnung einbezogenen Spiele ist das Vorzeichen ohne Klammer für die Summenbildung eingetragen (Summe I)

1 (+)	2 +	3 (-)	4 -	5 -
6 -	7 +	8 +	9 -	10 -
11 -	12 +	13 (-)	14 -	15 (-)

Zusätzlich wird nach der gleichen Methode eine Summe aus allen Spielen gebildet; das Vorzeichen in der Klammer deutet die Richtung der zusätzlich verrechneten Spiele an (Summe II). HERBST (1972) weist in einer weiteren Auswertung nach, daß sämtliche Spiele einen Zusammenhang mit den auf Grund der Notenabstände ermittelten und in zwei Gruppen nach linksverschobenen und rechtsverschobenen Skalen klassifizierten Gesamtskalen zeigen.

Empirische Ergebnisse der Auswertung der Notenskalen

Erste Schritte der Auswertung gelten der Absicherung der Berechnung der Werte für die Erfolgs-/Mißerfolgsorientierung.

Tabelle: III.III. - 1

Korrelationen der Skalengesamtwerte
(Summe I: 11 Spiele; Summe II: alle 15 Spiele)

	Summe I PVF	Summe I PFF	Summe II PVF	Summe II PFF	
Summe I PVF		.69	.98	.71	Hauptdiplom
Summe I PFF	.45		.73	.99	
Summe II PVF	.98	.47		.74	
Summe II PFF	.49	.97	.53		

Vordiplom

Methodisches Vorgehen: Die Korrelationen werden als Produkt-Moment-Korrelationen berechnet; im VD über die 43, im HD über die 29 Prüflinge hinweg. Sämtliche Korrelationen weichen mit $p < .01$ von $r = .00$ ab. Unter der Hauptdiagonalen sind die Korrelationen des VD, über der Hauptdiagonalen die des HD dargestellt.

Die beiden Summen weichen nur unwesentlich voneinander ab; das bestätigt indirekt die eindimensionale Struktur der Spiele. Über die Prüfung hinweg ist die Erfolgs-/Mißerfolgsorientierung im VD nur mäßig stabil. Die größere Stabilität im HD ist allerdings statistisch nicht abzusichern.

Eine andere Form der Stabilität kollektiver Art ist an den Parametern der Verteilungen der beiden Summen zu kontrollieren.

Tabelle: III.III. - 2

Mittelwertverschiebungen der Skalengesamtwerte Summe I und Summe II

Summe I	PVF		PFF		Wahrscheinlichkeit rein zuf. Mittelwertunterschiede
	arithm. Mittel	Varianz	arithm. Mittel	Varianz	
Vordiplom	17.44	8.32	16.69	9.56	n.s.
Hauptdiplom	16.29	11.46	15.93	11.17	n.s.
Wahrscheinlichkeit rein zuf. Mittelwertunterschiede	n.s.		n.s.		
Summe II					
Vordiplom	14.78	17.11	13.78	17.79	n.s.
Hauptdiplom	13.14	19.13	12.81	20.60	n.s.
Wahrscheinlichkeit rein zuf. Mittelwertunterschiede	n.s.		n.s.		

Methodisches Vorgehen: Die Höhe der Summen ergibt sich daraus, daß jeweils zu den individuellen Summen ein Wert von 20 addiert wird, um negative Summen zu verhindern. Die Mittelwertunterschiede werden zwischen den

beiden Prüfungsarten mit einem t-Test für unabhängige Stichproben nach Überprüfung der Homogenität der Varianzen getestet; zwischen den beiden Fragebögen erfolgt die Testung mit einem t-Test für abhängige Stichproben. Im VD sind 43 Prüflinge, im HD 29 Prüflinge Grundlage der Berechnung. Höhere Mittelwerte deuten eine Linksverschiebung (Mißerfolgsorientierung) der Skala an.

Die Unterschiede der Mittelwerte für VD und HD weisen zwar in die durch die unterschiedliche Notenverteilung vorgezeichnete Richtung stärker ausgeprägter Erfolgsorientierung im HD; der Unterschied ist jedoch äußerst gering. Bei den deutlichen Unterschieden in den Notenverteilungen zwischen VD und HD zeigt das, daß die in den Spielen realisierten Notenskalen eine mehr überdauernde Urteilstendenz erfassen, die von Situationsbedingungen eher unabhängig ist. Auch das Anspruchsniveau als konkretes Leistungsziel müßte, wenn es die Beantwortung begründen würde, einen deutlichen Unterschied bewirken. Die gleichfalls geringe Änderung über die Prüfung hinweg bestätigt die Situationsunabhängigkeit.

Die inhaltliche Auswertung im Rahmen der Stufenanalyse geht mit SCHMIDT (1966) davon aus, daß den subjektiven Notenskalen (hier als Erfolgs-/Mißerfolgsorientierung interpretiert) eine Bedeutung in der Verarbeitung von Fremdbewertung in der Selbstbewertung zukommt. Dies müßte sich in einer Moderatorfunktion bei der Wirkung der Noten auf die Gesamtkonzepte äußern: Die Wirkung sollte bei unterschiedlichen subjektiven Notenskalen unterschiedlich sein. Das kann durch eine zweifaktorielle Varianzanalyse mit Auspartialisierung der entsprechenden Antworten im PVF geprüft werden.

Tabelle: III.III. - 3

Interaktion von Skalengesamtwerten und Gesamtnoten (HD: "Weite Stichprobe")

	Wahrscheinlichkeiten rein zufälliger Interaktionen (Kovarianzanalyse)			
	Summe I		Summe II	
	PVF	PFF	PVF	PFF
Vordiplom				
Ges.Note x Summe gegen PFF-SK.GESAMT (PVF auspart.)	n.s.	n.s.	(=.08)	<.05
Ges.Note x Summe gegen PFF-SK.ZUKUNFT (PVF auspart.)	n.s.	n.s.	n.s.	n.s.
Hauptdiplom				
Ges.Note x Summe gegen PFF-SK.GESAMT (PVF auspart.)	n.s.	n.s.	n.s.	n.s.
Ges.Note x Summe gegen PFF-SK.ZUKUNFT (PVF auspart.)	n.s.	n.s.	n.s.	n.s.

Methodisches Vorgehen: Die Gesamtnoten werden in die Stufen "1" und "schlechter als 1" aufgeteilt; die Summen für die Notenskalen werden am Median dichotomisiert. Mit diesen beiden Faktoren als unabhängigen Variablen und jeweils einem der beiden Gesamtkonzepte, aus dem die Antworten im PVF auspartialisiert sind, als abhängiger Variabler wird eine zweifache Varianzanalyse berechnet. Die Berechnungen werden für die 43 Prüflinge des VD und die 46 der "Weiten Stichprobe" des HD, deren effektive Zahl wegen fehlender Antworten um einige Prüflinge geringer ist, durchgeführt.

Bei den Ergebnissen darf nicht verwundern, daß die Effekte für die beiden Summenwerte der subjektiven Notenskalen nicht übereinstimmen, obwohl diese eine

beinahe perfekte Korrelation aufweisen (s. Tabelle: III.III. - 1). In die Varianzanalyse gehen nur die am Median dichotomisierten Summen ein. Durch die Dichotomisierung sinkt der Zusammenhang zwischen den beiden Summen auf eine Korrelation um .80 ab. Diese Korrelation aber läßt Unterschiede zu.

Nur im VD hängt die Wirkung der Gesamtnoten auf das gegenwärtige Gesamtkonzept von der Erfolgs-/Mißerfolgsorientierung der Prüflinge ab. Diesem Effekt ist insofern eine große Bedeutung zuzumessen, als in der Varianzanalyse der entsprechende Haupteffekt der Gesamtnoten, der als partielle Korrelation signifikant ist (s. Tabelle: III.II. - 3), nicht nachzuweisen ist. Da die Erfolgs-/Mißerfolgsorientierung die Moderatorwirkung als Tendenz auch aufweist, wenn sie vor der Prüfung gemessen wird, ist eine kausale "Mitwirkung" dieser personalen Tendenz wahrscheinlich. Sie stellt im Sinne der Normwerte eine Verfälschung dar, weil die Normwerte eine universelle, von personalen Bedingungen unabhängige Wirkung der Noten fordern. Die Beschränkung der Verfälschung auf das gegenwartsbezogene Gesamtkonzept verschärft diese Bewertung eher, weil die Normwerte eine Wirkung auf dieses in den Vordergrund rückten.

Eine weitere Bedeutung kann die Erfolgs-/Mißerfolgsorientierung als mögliche Alternativreaktion auf das Ergebnis der Prüfung gewinnen. Die geringfügigen Änderungen der Mittelwerte der beiden Summen schließen nicht aus, daß Änderungen in Abhängigkeit von den erteilten Noten erfolgen. Die Korrelationen der Summen im PVF und PFF zeigen, daß Änderungen in der Erfolgs-/Mißerfolgsorientierung besonders im VD ein großes Ausmaß haben (Tabelle: III.III. - 1).

Tabelle: III.III. - 4

Wirkung der Gesamtnoten auf die Änderung der Skalengesamtwerte

	Wahrsch.zufälligen Zusammenhangs Ges.Note-Summe PFF (Kovarianzanalyse)	Part. Korrelation Ges.Note/Summe PFF (PVF auspartialis.)	Wahrsch. zuf.Abw. von $r_{part.}=.00$
Summe I			
Vordiplom	(=.15)	.19	n.s.
Hauptdiplom ("Weite Stichprobe")	<.05	.35	<.05
Summe II			
Vordiplom	(=.14)	.18	n.s.
Hauptdiplom ("Weite Stichprobe")	(=.07)	.27	(=.07)

Methodisches Vorgehen: In der Kovarianzanalyse werden die Gesamtnoten mit den Stufen "1" und "schlechter als 1" als Faktor, die Summen für die Notenskalen des PFF als abhängige Variable, aus der die Summen des PVF auspartialisiert sind, in einer einfachen Varianzanalyse verrechnet. Die Berechnung der partiellen Korrelation erfolgt analog. Die Zahl der Prüflinge beträgt im VD 43; im HD ist wegen der Streuung der Gesamtnoten nur die "Weite Stichprobe" mit 46 Prüflingen auswertbar.

Im VD sind, trotz geringerer Korrelationen der Summen zwischen PVF und PFF, nur Tendenzen zur Abhängigkeit der Erfolgs-/Mißerfolgsorientierung von der erhaltenen Gesamtnote vorhanden: Die Erfolgsorientierung nimmt bei besseren Noten etwas zu. Dagegen ist im HD bei größerer Stabilität der Summen diese Tendenz statistisch bedeutsam. Sie könnte eine wirksame Alternativreaktion darstellen; die Verarbeitung der Notenrückmel-

dung im HD erfolgte dann durch eine Umorientierung gegenüber dem Notensystem und erübrigte eine Änderung des Selbstkonzepts. (Eine Ausnahme allerdings bildet das zukunftsbezogene Selbstkonzept, das auch im HD von den Noten beeinflußt wird: Möglicherweise stellt die mehr affektbezogene Umorientierung nur eine Alternative für rationale Prozesse wie das gegenwärtige Selbstkonzept dar, während sie mit anderen affektbezogenen Verarbeitungen der Noten wie Motivationsänderungen und dem"affektnäheren" zukunftsbezogenen Gesamtkonzept eine Einheit bildet. Zur Abklärung reichen die Ergebnisse dieser Untersuchung nicht aus.)

Abschließend soll eine deskriptive Auswertung der Struktur der subjektiven Notenskalen versucht werden. In Abweichung von dem bei SCHMIDT (1966) und HERBST (1972) gewählten Vorgehen, die kollektiven Skalen über die Auswertung der individuellen Skalen zu bestimmen, wird versucht, von den Antworthäufigkeiten der Gesamtgruppe direkt solche kollektiven Skalen abzuleiten. Die Auswertung baut auf den Spielen auf, bei denen von der Gruppe der Prüflinge (VD und HD getrennt) statistisch bedeutsam ein Paar häufiger gewählt wird als das andere. Die Spiele werden durch Berechnung von CHI-Quadrat der Abweichung von der Gleichverteilung mit einer Irrtumswahrscheinlichkeit von $p < .05$ identifiziert. Nach diesem Vorgehen sind für das VD im PVF 10 der Spiele, im PFF noch 6 der Spiele einseitig beantwortet. Für das HD betragen die entsprechenden Häufigkeiten: im PVF 9 Spiele, im PFF 8 Spiele.

Die Konstruktion der kollektiven Skalen aus den Spielen erfolgt, indem die Abstände zwischen den Noten so geordnet werden, daß sie die Präferenzwahlen widerspiegeln (BECKER & SIEGEL 1962). Das Verfahren schließt unmittelbar an das im Anfang erwähnte Beispiel der Bestimmung der Abstände zwischen den Notenstufen an. Nach diesem Verfahren lassen sich die Präferenzwahlen aus

der Ordnung der Abstände ableiten. Die Abstände stellen lediglich Einheiten dar, die zueinander in der Relation "kleiner", "gleich" und "größer" stehen. Es wird dabei jene Darstellung gewählt, die mit der geringsten Zahl von Einheiten der Abstände die Präferenzen abbildet. Für die vier Bedingungen ergeben sich die folgenden Skalen.

Tabelle: III.III. - 5

Kollektive Notenskalen

Vordiplom

PVF 1 + 2 + + + 3 + + 4 + + + + 5

PFF 1 + 2 + + 3 + 4 + + 5

Hauptdiplom

PVF 1 + 2 + + 3 + 4 + + 5

PFF 1 + 2 + + 3 + 4 + + 5

Methodisches Vorgehen: Die Anzahl der "+" zwischen den Notenstufen deutet die Ordnungsrelation der einzelnen Abstände an, die für eine Herleitung der Richtung der Präferenzwahlen ausreichen. Es sind nur Ordnungsbeziehungen zwischen einzelnen Abständen notwendig, nicht solche zwischen Kombinationen von Abständen (higher ordered metric scale).

Bei der Auswertung treten keine einander widersprechenden Präferenzen auf, wie sie von HERBST (1972) für die individuellen Notenskalen berichtet werden. Drei der vier erhaltenen kollektiven Notenskalen sind identisch. Aber auch die Skala für VD/PVF weist mit den anderen drei Skalen zusammen eine ähnliche Grobstruktur auf: Die Noten "1" und "2" zeigen einen geringeren Nützlichkeitsunterschied, ebenso die Noten "3" und "4". Demgegenüber hat die Note "5" einen größeren Abstand; eine ähnliche Trennung findet zwischen der

Gruppe der Noten "1" und "2" und den übrigen Noten statt. Die Notenskalen sind eindeutig auf die Meidung von Mißerfolg zentriert. Die Meidung von Mißerfolg hat zwei Stufen und zeigt darin, daß die Notenskala nicht eine einheitliche Dimension darstellt. Sie besteht zum einen in der Meidung der "5", die besonders deutlich bei VD/PVF von den übrigen Noten abgesetzt ist. Die restliche Skala ohne die "5" wiederum zerfällt in die beiden "guten" Noten und die beiden "schlechten" Noten. Trotz der unterschiedlichen Zahl der extrem beantworteten Spiele im PVF und PFF sind die Skalen, die den Wahlen zugrunde gelegt werden können, nicht unterschiedlich. Eine wesentliche Veränderung, die auch bei den Summen nicht nachgewiesen werden konnte, findet über die Prüfung hinweg nicht statt. Das Ergebnis bestätigt die Vermutung, daß die subjektiven Notenskalen keine Gleichabständigkeit der Notenstufen verwirklichen. Es macht darüber hinaus deutlich, daß die Prüfungen von den Prüflingen eindeutig mißerfolgszentriert erlebt werden. Die geringe Wirkung der Gesamtnoten, die im wesentlichen zwischen "1" und "2" differenzieren, ist als Folge der geringen subjektiven Bedeutung dieses Unterschiedes zu sehen.

Zusammenfassung (III.III.1.)

Die subjektiven Notenskalen der Prüflinge können als Moderatorvariablen oder als alternative abhängige Variablen in die Wirkung der Gesamtnoten auf die Gesamtkonzepte eingreifen. Das Verfahren zur Ermittlung der subjektiven Notenskalen baut auf der Nutzenskalierung auf. Einige Untersuchungen legen nahe, daß mit dem Verfahren das Anspruchsniveau angesprochen wird. Doch Ergebnisse anderer Untersuchungen bestätigen das nicht. Auf einer RASCH-Skalierung fußend wird die Rechts-/Linksverschiebung der Notenskalen als die wesentliche Dimension der Antworten bestimmt, deren

psychologische Grundlage als Erfolgs-/Mißerfolgsorientierung gedeutet wird. Zur Kennzeichnung der Dimension werden aus den einzelnen Spielen der Nutzenskalierung durch einfache Summierung von 11 bzw. 15 Spielen im PVF und PFF jeweils zwei Gesamtwerte berechnet. Die Korrelationen dieser Gesamtwerte zwischen PVF und PFF sind im VD nur mäßig hoch, im HD (nicht signifikant) höher. Die Mittelwerte der Gesamtwerteverteilungen sind geringfügig zwischen den beiden Prüfungsarten und den beiden Befragungen (PVF und PFF) unterschiedlich. Die Gesamtwerte messen somit eine eher situativ überdauernde Orientierung.

Die Notenskala als Moderatorvariable wird über die Interaktion mit den Gesamtwerten in einer zweifachen Varianzanalyse mit Kovariatenkontrolle erfaßt. Die Interaktion ist bei einem der beiden Gesamtwerte für das gegenwärtige Gesamtkonzept im VD statistisch bedeutsam. Das zeigt eine Wirkung an, die nicht im Sinne der Normwerte ausfällt.

Als Alternativreaktion kommt eine Änderung der Notenskala in Abhängigkeit von den erteilten Gesamtnoten in Betracht: Sie ist für das HD in statistisch bedeutsamem Ausmaß festzustellen. Eine solche Änderung der subjektiven Notenskalen könnte eine Änderung der Selbstkonzepte erübrigen.

Die zusätzliche deskriptive Auswertung der Struktur der kollektiven subjektiven Notenskalen zeigt, daß die Prüflinge in ihrer Gesamtheit mißerfolgsorientiert sind. Diese Mißerfolgsorientierung findet ihren Ausdruck in einer Meidung der "5" und geringen Differenzierung der "1" und der "2" und macht von daher gesehen die geringe Wirkung der (Gesamt-) Noten verständlich.

III.III.2.: Das Anspruchsniveau als Grundlage erlebten Erfolgs und Mißerfolgs

In allen grundlegenden Arbeiten zum Anspruchsniveau wird betont, daß das Anspruchsniveau der Person bestimmt, wann eine Leistung als Erfolg oder Mißerfolg erlebt wird (FRANK 1941; HOPPE 1931; LEWIN et al. 1944). Für die Wirkung der Noten kann vermutet werden, daß nicht die objektive Note, sondern die subjektiv als Erfolg oder Mißerfolg erlebte Note die dynamische Grundlage einer Änderung von Selbstkonzept und Motivation ist. Die Auswertung der Arbeiten zu Änderungen des Selbstkonzepts (Kapitel III.II.A.1.) läßt annehmen, daß die Wirkungen von Erfolg und Mißerfolg qualitativ unterschiedlich sind.

Bei der Verwendung des Anspruchsniveaus zur Bestimmung erlebten Erfolgs und Mißerfolgs ergibt sich ein methodisches Problem. Es setzt an der Meßbarkeit des Anspruchsniveaus durch Auskunftgeben an: HOPPE (1931) sieht das Anspruchsniveau als komplexe, hierarchisch organisierte psychische Struktur, die nicht unbedingt durch einfaches Abfragen zu ermitteln ist. In bestimmten Aspekten nimmt das Anspruchsniveau eher die Stellung eines hypothetischen Konstrukts ein, wenn HOPPE empfiehlt, es aus dem Erleben von Erfolg oder Mißerfolg zu erschließen. Auch die theoretische Verwendung bei LEWIN et al. (1944) in einer Vorwegnahme der Theorie der Leistungsmotivation läßt an der Operationalisierung durch Abfrage Zweifel aufkommen. Die Entwicklung einer eigenständigen Meßtradition erfolgte zudem in einer Richtung, die für die hier beabsichtigte Verwendung weniger geeignet erscheint. Das Anspruchsniveau wird als eine Eigenschaft der Person verstanden, die über die Anspruchsniveausetzung bei einfachen Aufgaben ermittelt wird; bewährt haben sich für solche Messungen insbesondere Diskrepanzwerte zur vorhergehenden Leistung (HECKHAUSEN 1965). Hier aber

soll das Anspruchsniveau die konkrete Zielsetzung als angestrebte Leistung in der einzelnen Prüfung erfassen.

Stellt man die prinzipiellen Bedenken gegen eine Abfrage als konkrete Zielsetzung, die auch von HOPPE (1931) nicht durchgehend geteilt werden, zurück, dann steht als weiteres Problem das der Ebene an, auf der das Anspruchsniveau erfragt werden soll. Entsprechend der hierarchischen Struktur ist beispielsweise eine Abfrage als Wunsch oder als Erwartung möglich. Es hat sich herausgestellt, daß die Abfrage als Erwartung zu realistischeren Einschätzungen führt (HOLT 1946; LOCKE & BRYAN 1968). Dabei kann allerdings bezweifelt werden, ob mit den unterschiedlichen Abfragen eine einheitliche Struktur angesprochen wird; IRWIN (1944) trägt dem durch eine weitere Begriffsdifferenzierung (level of expectation) Rechnung, die sich aber nicht durchgesetzt hat (sie wird auch bei LEWIN et al.(1944) verwendet). Nach LOCKE & BRYAN (1968) stimmen die unterschiedlichen Abfragen gut miteinander überein. Zur Bestimmung von Erfolg oder Mißerfolg liegt eine Verwendung der realistischeren Erwartung nahe.

Es sind bisher wenig Erfahrungen mit der Eignung solcher konkreten Zielsetzungen zur Bestimmung von subjektiv erlebtem Erfolg und Mißerfolg vorhanden. Diese Tatsache für sich genommen läßt weitere Zweifel an dem beabsichtigten Vorgehen aufkommen. Ergebnisse der Verwendung von Anspruchsniveaumessungen in anderen Kontexten bestärken die Zweifel eher. Zwar bestätigt WORELL (1959) eine relativ hohe zeitliche Stabilität einer angestrebten Note über vier Jahre hinweg ($r = .60$), doch werden auf der anderen Seite auch bedeutende zeitliche Schwankungen berichtet (HOPPE 1931). Die (soziale) Dynamik der Abfragesituation determiniert das berichtete Anspruchsniveau (FRANK 1941). Dies ist besonders gewichtig im Zusammenhang mit Ergeb-

nissen von HOLT (1946), der bei Leistungen mit bedeutsamer Ich-Beteiligung, wie sie Prüfungen sicher darstellen, dem Anspruchsniveau eher die Bedeutung einer Ich-Abwehr zumißt. Nachweise der Abhängigkeit des Anspruchsniveaus von bestimmten personal stabilen Eigenschaften (wie der Ausprägung der überdauernden Motivation; WEINER 1976) oder die damit vermutlich vermittelte Abhängigkeit der Realistik des Anspruchsniveaus von der absoluten Höhe der Studienleistungen (MURSTEIN 1965; WORELL 1959) sind ein weiterer möglicher Einwand, wenngleich solche Tendenzen auch valide Ursache subjektiven Erfolgserlebens sein dürften (zumindest aber wird dadurch die getrennte Verrechnung der einzelnen Prüfungen fragwürdig).

Die Verwendung der für die einzelnen Fächer und die Prüfung insgesamt als Notenerwartung (PVF-I.NOTENERW.GESAMT; PVF - F 8. NOTENERWARTUNG) ermittelten Anspruchsniveaus zur Berechnung des subjektiven Erfolgs und Mißerfolgs kann durch zwei Auswertungsschritte vorbereitet werden.

Der subjektive Erfolg oder Mißerfolg wird als Differenz der erwarteten Note und der wirklichen Note berechnet. Diese Differenz ist gleichzeitig ein Maß für die Realistik des Anspruchsniveaus.

Tabelle: III.III. - 6

Realistik des Anspruchsniveaus
bei Gesamtnoten und Einzelnoten

	Prozent der Antworten			Anzahl der Prüflinge
	Erwartung übertroffen	Erwartung erreicht	Erwartung nicht erreicht	
Gesamtnoten				
VD	75	22	3	40
HD ("Weite Stichprobe")	34	63	3	32

Tabelle: III.III. - 6 (Fortsetzung)

Realistik des Anspruchsniveaus
bei Gesamtnoten und Einzelnoten

	Prozent der Antworten			Anzahl
	Erwartung übertroffen	Erwartung erreicht	Erwartung nicht erreicht	der Prüfungen
"Über-alle-Fächer"				
VD	60	27	13	282
VD (nur psychologische Fächer)	60	26	14	172
HD	46	49	5	145

Methodisches Vorgehen: Durch Fehlen der Frage PVF-I.NOTENERW.GESAMT in der Vorform I ist die Zahl der Prüflinge in der "Weiten Stichprobe" beim HD, die wegen der geringen Notenstreuung verwendet werden muß, gegenüber der Ausgangszahl von 46 reduziert. Im VD werden zusätzlich bei der Auswertung "Über-alle-Fächer" nur die psychologischen Prüfungen ausgewertet.

Den Ergebnissen ist zu entnehmen, daß im Mittel die Notenerwartungen deutlich unter den wirklich erreichten Noten liegen. Das könnte aus der Unkenntnis der Verteilung der wirklichen Noten und entsprechender Orientierung an vorgängig erfahrenen Verteilungen (VD: Schule; HD: VD) zu erklären sein. Eine alternative Erklärung schließt sich an HOLT (1946) an, der bei Leistungen mit hoher Ich-Beteiligung das Anspruchsniveau als im Dienste der Ich-Abwehr stehend sieht. Ähnlich machen FRANK (1941) und schon HOPPE (1931) auf die Möglichkeit aufmerksam, daß das Anspruchsniveau in bestimmten Situationen sehr tief gesetzt wird, um einen Mißerfolg unbedingt zu vermeiden. Die Ergebnisse der subjektiven Notenskalen in Teil 1 bestätigen, daß die Meidung von Mißerfolg als Orientierung dominiert. Die Richtung der Abweichung steht in einem auffälligen Ge-

gensatz zu amerikanischen Untersuchungen, die übereinstimmend eine Tendenz zur Überschätzung (auch bei Erwartungen) finden (LOCKE & BRYAN 1968; MURSTEIN 1965; PETZEL 1972). Das bestätigt die Bedeutsamkeit der situativen Einbettung, die unter diesen (Abfrage-)Bedingungen wohl eher die Abwehr in binnendynamischer Richtung aktiviert, während unter amerikanischen Verhältnissen die Abwehr in Richtung sozialer Selbstdarstellung ins Spiel kommt. Beide Erklärungen der Abweichungen schränken die Brauchbarkeit der Erwartungen zur Ableitung des subjektiven Erfolgs und Mißerfolgs ein: Die falschen Vorannahmen über die Notenverteilungen sind nach der Prüfung durch die Prüflinge selbst korrigierbar; die Binnendynamik der Abwehr kann konstante personale Tendenzen motivationaler Art vermehrt wirksam werden lassen (s. oben).

Die empirische Verteilung der Werte für den erlebten Erfolg und Mißerfolg läßt eine getrennte Auswertung im Hinblick auf qualitative Unterschiede von Erfolg und Mißerfolg, wie sie anfänglich beabsichtigt war, nicht zu.

Für die auf die einzelnen Prüfungen bezogenen Anspruchsniveaus kann durch eine Korrelation zwischen Paaren von Einzelfächern die fachspezifische Ausbildung der Erwartungen, die gleichzeitig ein Maß für die Bedeutung personal stabiler Tendenzen darstellt, überprüft werden. Im VD beträgt die über alle Paare einzelner Fächer hinweg gemittelte Produkt-Moment-Korrelation .28 mit einem Streubereich von .01 bis .65. (Dabei wird das Mittel als arithmetisches Mittel berechnet; die einzelnen Korrelationen werden nach der Zahl der Prüflinge gewichtet.) Die entsprechenden Werte für das HD belaufen sich auf eine mittlere Korrelation von .42 mit einem Streubereich von .20 bis .64. Die Zusammenhänge beim VD sprechen für eine dominant situationsspezifische Ausrichtung der Erwartungen, wäh-

rend im HD schon mehr personale Konstanz vorhanden ist. Der überwiegende Teil der Varianz dürfte, selbst bei der Annahme nicht besonders hoher Zuverlässigkeit, situationsspezifisch determiniert sein. Besonders beim HD, aber auch beim VD ist jedoch ein Einfluß personal stabiler Tendenzen möglich.

Die Auswertung der Wirkung des erlebten Erfolgs und Mißerfolgs (als Differenz von Noten und Erwartungen) soll in einem dimensionalen Ansatz erfolgen, indem der erhaltene Wert für Erfolg und Mißerfolg wie die ursprünglichen Noten verrechnet wird. Wenn nicht die Höhe der Noten, sondern das Ausmaß des erlebten Erfolgs und Mißerfolgs die eigentliche Quelle der Wirkungen darstellt, müssen die partiellen Korrelationen zwischen Erfolg/Mißerfolg und den thematischen Variablen enger sein als die zwischen den Noten und diesen Variablen. Die Auswertung beschränkt sich deshalb auf einen Vergleich der partiellen Korrelationen.

Tabelle: III.III. - 7

Vergleich der Wirkung der Noten und des erlebten Erfolgs/Mißerfolgs auf Selbstkonzept und Motivation (Gesamtkonzepte und "Über-alle-Fächer")

	Partielle Korrelationen			
	Vordiplom		Hauptdiplom	
	Noten/ PFF (PVF auspart.)	Erfolg-Mißerfolg/ PFF (PVF auspart.)	Noten/ PFF (PVF auspart.)	Erfolg-Mißerfolg/ PFF (PVF auspart.)
Gesamtkonzepte				
SK.GESAMT	+.31	+.22	+.06	-.03
SK.ZUKUNFT	+.34	+.23	+.40	+.15
"Über-alle-Fächer"				
SK.FÄCHER	+.29	+.27	+.17	+.07
MOT.W.A.	+.15	+.13	+.15	-.15
MOT.BEL.	+.17	+.12	+.24	-.07
nur psych. Fächer				
SK.FÄCHER	+.39	+.35		
MOT.W.A.	+.13	+.11		
MOT.BEL.	+.09	+.07		

Methodisches Vorgehen: Die Berechnungen erfolgen wie die für Tabelle: III.II. - 2/3. Bei den Auswertungen der Gesamtkonzepte ist wegen fehlender Antworten die Zahl der einbeziehbaren Prüflinge gegenüber der Ausgangsstichprobe auf 39 beim VD und 31 beim HD (wegen der Notenstreuung nur "Weite Stichprobe") verringert. Die Zahl der Prüfungen beträgt wie immer um 280 im VD und 145 im HD. Zusätzlich wird zum Vergleich mit der Wirkungsanalyse der Noten im VD die Auswertung allein für die psychologischen Prüfungen vorgenommen; dort beträgt die Zahl der Prüfungen um 170. Eine positive Korrelation bedeutet, daß mit zunehmendem Erfolg die Wirkung in die Richtung größerer Motivation und höheren Selbstkonzepts zielt.

Unter keiner Auswertungsbedingung ist eine Wirkung des erlebten Erfolgs/Mißerfolgs festzustellen, die diejenige der objektiven Noten überträfe. Es kann damit ausgeschlossen werden, daß die Wirkung der Noten wesentlich über den subjektiven Erfolg/Mißerfolg vermittelt wird, wenn dieser aus dem Anspruchsniveau vor der Prüfung bestimmt wird. Der positive Nachweis einer im Sinne der Normwerte verfälschenden Wirkung ist nicht gelungen.

Das Ergebnis entspricht den Erkenntnissen aus der Faktorenanalyse (s. Kapitel III.II.B.6.), die eine durchgehende Unabhängigkeit der auf die Prüfung gerichteten Aspekte des Erlebens vor der Prüfung von denen nach der Prüfung fand. So kann vermutet werden, daß das Anspruchsniveau als Ich-Abwehr allein der Bewältigung der besonderen Situation vor der Prüfung dient und deshalb wenig funktionale Beziehungen zur Verarbeitung der Erfahrungen in den Prüfungen hat. Eine weitere Bedingung der Begrenzung der Wirkung des an den Erwartungen gemessenen Erfolgs/Mißerfolgs ist auch in den Untersuchungen zur Attribuierung zu finden: Ergebnisse einer Prüfung werden dann vermehrt äußeren Faktoren zu-

geschrieben, wenn sie nicht den Erwartungen entsprechen (SIMON & FEATHER 1973); das wird im folgenden Teil untersucht.

Zusammenfassung (III.III.2.)

Der durch das individuelle Anspruchsniveau bestimmte subjektive Erfolg und Mißerfolg wird als die eigentliche Quelle der Wirkung vermutet. Er wird über die Differenz von Notenerwartungen und Noten berechnet.

Vorliegende Arbeiten lassen an der Möglichkeit einer direkten Abfrage des Anspruchsniveaus zweifeln. Wenn überhaupt, scheint eine Abfrage über die konkreten Erwartungen sinnvoll, da sie zu realistischen Einschätzungen führen soll. Eine Auswertung der Erwartungen zeigt, daß sie dominant in die Richtung von Unterschätzungen weisen. Das mindert ihre Brauchbarkeit zur Ableitung von subjektivem Erfolg und Mißerfolg und macht eine abgetrennte Auswertung für den erlebten Mißerfolg nicht möglich. Die Korrelationen zwischen den Erwartungen in den einzelnen Fächern deuten auf eine überwiegend situativ bedingte Varianz hin, wenngleich besonders beim HD personal stabile Tendenzen erheblich sind.

Der Vergleich der partiellen Korrelationen, bei denen statt der Noten der erlebte Erfolg/Mißerfolg als unabhängige Variable verwendet wird, ergibt, daß keine der Korrelationen diejenige mit den Noten übertrifft. Der subjektiv erlebte Erfolg und Mißerfolg stellt diesbezüglich deshalb keine Verfälschung im Sinne der Normwerte dar.

III.III.3.: Die erlebte Selbstverantwortlichkeit für die Noten

Der Begriff der Selbstverantwortlichkeit wird in diesem Teil mehr beschreibend für einen Prozeß verwendet, der allgemeiner als Attribution (Ursachenzuschrei-

bung) bezeichnet wird. Schon HOPPE (1931) merkt an, daß Erfolg oder Mißerfolg in einer Handlung von den Trägern der Handlung unterschiedlich verarbeitet werden kann; solche Unterschiede gehen u.a. auf das "Abschieben der Verantwortung" zurück. Im Anschluß an HEIDER hat sich eine Theorie entwickelt (MOSCOVICI (1975) bestreitet, daß es sich um mehr als eine "Sicht" handele), die auch zur Differenzierung der Leistungsmotivationsforschung beigetragen hat (WEINER 1976). Ähnlich wie beim Anspruchsniveau ist das ursprünglich sowohl personal als auch situativ gebunden gedachte Konstrukt in weitem Umfang auf ein differentielles Konstrukt reduziert worden, wie beispielsweise bei der Unterscheidung von internalem und externalem "locus of control" nach ROTTER oder bei der Selbstverantwortlichkeit (MEYER 1973). Demgegenüber soll der Begriff hier nicht auf eine solche Verengung festgelegt werden.

Das schließt nicht aus, auch solche Erfahrungen, die mit personalistisch verstandenen Konstrukten gesammelt wurden, zu integrieren. Erkenntnisse aus diesem Bereich sind Grund zur Annahme, daß die erlebte Selbstverantwortlichkeit an der Wirkung der Noten auf Selbstkonzept und Motivation beteiligt ist. So kann nachgewiesen werden, daß Rückmeldungen über Erfolg und Mißerfolg bei Personen mit unterschiedlichem Stil der erlebten Verursachung (internal versus external) eine unterschiedliche Wirkung haben (DAVIS & DAVIS 1972; WEIS & SHERMAN 1973), wenngleich die Unterschiede eher bei der Reaktion auf Mißerfolg deutlich werden, der nach den Ergebnissen von Teil 2 hier selten sein dürfte. WEINER (1976) belegt durch mehrere Untersuchungen, daß die Attribution in die Auswirkungen motivierten Verhaltens eingreift. Gehen diese Hinweise mehr in die Richtung einer "Mitwirkung" bei dem Einfluß auf die Motivation, deuten andere darauf hin, daß die Attribution gleichfalls an der Wirkung auf das Selbstkonzept

beteiligt sein könnte: Bei internaler Zuschreibung wird Information besser aufgenommen und verarbeitet; ein Nachweis von unterschiedlicher Berücksichtigung der Rückmeldung in Prüfungen bei folgenden Erwartungen mißlingt allerdings (WOLFE 1972). Prozesse der Zuschreibung sind an der Aktualgenese von Selbstkonzepten beteiligt (DUVAL & WICKLUND 1972). Die erlebte Selbstverantwortlichkeit kann mit der Glaubwürdigkeit in Verbindung gebracht werden, die sich als eine wesentliche situative Bedingung der Wirkung auf das Selbstkonzept erwies.

Die erlebte Selbstverantwortlichkeit für die Note wird über die Bewertung der Note durch den Prüfling erfaßt, sowohl fachbezogen (PFF-F 10.WIRKL.LEISTUNG und PFF-F 11.STUDIENLEISTUNG) als auch auf die Gesamtnote bezogen (PFF-IV.LEISTUNG.GESAMT). Die Bewertung der Noten der mündlichen Prüfungen ist in zwei Bezugssystemen möglich. Die Noten können zum einen vom Prüfling an dem Eindruck gemessen werden, den er vom Verlauf der Prüfung und der Begründung der Note gewonnen hat. Zum anderen kann auch seine von ihm selbst eingeschätzte bisherige Leistung im Studium als ein Bezugssystem dienen.

Diese "Bewertungen von Bewertungen" weichen insofern von den üblichen Operationalisierungen von Zuschreibungen ab (z.B. WEINER 1976), als die Bezugssysteme, auf Grund derer eine Leistung der eigenen Person zugeschrieben wird oder nicht, explizit vorgegeben werden. Im Anschluß an die einleitenden Bemerkungen dieses Kapitels ist zu betonen, daß die erlebte Selbstverantwortlichkeit nicht einen kausalen Status gegenüber den Wirkungen der Noten beanspruchen kann, sondern nur die Kennzeichnung eines Änderungsprozesses ermöglicht, der nicht auf die Noten zurückgeht. In dieser Absicht soll sie als Moderatorvariable verwendet werden.

Wie bisher sollen einige Eigenschaften der Antworten vorab analysiert werden; an erster Stelle sind das die Zusammenhänge zwischen den Antworten.

Tabelle: III.III. - 8

Korrelationen der Fragen zur Selbstverantwortlichkeit

	Vordiplom			Hauptdiplom		
	Kor-rel.	Streu-bereich von	bis	Kor-rel.	Streu-bereich von	bis
Korrelationen zwischen den Fragen						
PFF - F 10./PFF - F 11.						
"Fach-Prüfer-Kombinationen" (mittlere Korrelation)	.76	.39	.84	.48	.07	.70
"Über-alle-Fächer"	.62			.37		
PFF - F 10./PFF-IV.						
Einzelfächer (mittlere Korrelation)	.45	.36	.53	.12	-.05	.30
PFF - F 11./PFF-IV.						
Einzelfächer (mittlere Korrelation)	.40	.04	.61	.25	.08	.36
Korrelationen innerhalb der Fragen						
Paare von Einzelfächern (mittlere Korrelation)						
PFF - F 10.WIRKL.LEIST.	.22	-.02	.43	.17	-.12	.58
PFF - F 11.STUDIENLEIST.	.22	-.12	.47	.11	-.08	.32

Methodisches Vorgehen: Alle Korrelationen werden als Produkt-Moment-Korrelationen berechnet. Bei Mittelungen der Korrelationen aus den "Fach-Prüfer-Kombinationen" werden die Werte wegen der großen Streuung der Korrelationen z-transformiert und nach der Zahl der Freiheitsgrade gewichtet. Bei Mittelungen über die Einzelfächer wird nur im VD nach der Zahl der Prüflinge

gewichtet; eine z-Transformation unterbleibt dort wegen nur geringfügiger Auswirkungen.

Der erste Teil der Tabelle faßt Korrelationen zusammen, die zwischen den einzelnen Fragen berechnet werden. Im zweiten Teil werden die Korrelationen für die beiden einzelfachbezogenen Fragen zwischen den Antworten zu je zwei Fächern berechnet und über alle so gebildeten Paare hinweg (im VD gewichtet) gemittelt.

Die Zusammenhänge zwischen den beiden fachbezogenen Fragen sind ein Maß dafür, ob die Antworten mehr globale Wertungen der Noten darstellen, oder ob es den Fragen gelingt, unterschiedliche Bezugssysteme anzusprechen. Im HD sind die Antworten nur mäßig voneinander abhängig, während im VD eine größere Abhängigkeit besteht. Dieser Unterschied ist mit dem größeren eigenständigen Gewicht des Studiums im HD zu erklären (ALVES & LODDERS 1974), der sich auch in der kompetenzorientierten Ausprägung des Selbstkonzepts im HD beweist.

Die Korrelationen innerhalb der beiden fachbezogenen Fragen sind ein Maß für die situationsspezifische Differenzierung der Antworten. Die geringen mittleren Korrelationen zeigen, daß die wesentliche Varianz situationsspezifisch determiniert ist. Doch die Korrelation der fachbezogenen Fragen mit der auf die Gesamtnote bezogenen Frage (Zeile 3 und 4 der Tabelle: III.III. - 8) macht für das VD deutlich, daß dort eine persönlichkeitsspezifische Bedingung der Varianz in etwas größerem Ausmaß vorliegt.

Für die Verrechnung der Antworten zur Selbstverantwortlichkeit wird durch die Unterscheidung internaler und externaler Zuschreibung eine rein formale Aufteilung in Ankreuzungen der mittleren Skalenposition ("zutreffend") und die Gruppe der übrigen Ankreuzungen nahegelegt. Eine Alternative wäre eine dimensionale Auswertung, die aber weniger präzise dem Begriff der Selbst-

verantwortlichkeit zuzuordnen ist. Die empirische Verteilung der Antworten macht aus dieser Unterscheidung weitgehend ein Scheinproblem.

Tabelle: III.III. - 9

Verteilung der Antworten zur Selbstverantwortlichkeit

	Prozentuale Verteilung (eher) zu gut	zu-treffend	(eher) zu schlecht	Zahl d. Prüf./ Prüfl.
Gesamtwerte				
Vordiplom				
PFF-IV.LEISTUNG.GESAMT	50	33	17	42
Hauptdiplom				
PFF-IV.LEISTUNG.GESAMT ("Weite Stichprobe")	69	31	-	35
"Über-alle-Fächer"				
Vordiplom				
PFF-F 10.WIRKL.LEISTUNG	34	52	14	281
PFF-F 11.STUDIENLEIST.	42	43	15	
nur psychol. Fächer				
PFF-F 10.WIRKL.LEISTUNG	43	46	11	170
PFF-F 11.STUDIENLEIST.	49	39	12	
Hauptdiplom				
PFF-F 10.WIRKL.LEISTUNG	39	56	5	145
PFF-F 11.STUDIENLEIST.	44	50	6	

Methodisches Vorgehen: Die beiden extremen Positionen der 5-stufigen Antwortskala sind jeweils in einer Kategorie zusammengefaßt. Die Anzahl der Prüflinge für die Frage PFF-IV.LEISTUNG.GESAMT im HD liegt weit unter der von 46 Prüflingen, weil in der Vorform I diese Frage fehlt.

Die Antworten, die eine Fremdzuschreibung andeuten, liegen überwiegend in positiver Richtung: Die Prüf-

linge sehen die Note eher als Überschätzung ihrer wirklichen Leistungen. Dies entspricht dem im vorangehenden Teil festgestellten häufigen Übertreffen der Erwartungen. Wegen dieser einseitigen Verteilung kann eine kategoriale Aufteilung (in die Wahl der mittleren Skalenposition und die Wahl aller übrigen) von einer dimensionalen Aufteilung (z.B. durch Dichotomisierung am Median) nicht klar unterschieden werden. Die Auswertung wird, wie es der theoretische Hintergrund nahelegt, bevorzugt in kategorialer Aufteilung durchgeführt (Korrelationen dagegen werden dimensional bestimmt).

Die Überprüfung der Moderatorfunktion erlebter Selbstverantwortlichkeit erfolgt in einer zweifachen Varianzanalyse mit Kovariatenkontrolle über die Interaktion von Selbstverantwortlichkeit und Noten als Faktoren. Die Noten werden wie üblich zu zwei Stufen ("1" und "schlechter als 1") zusammengefaßt; die Selbstverantwortlichkeit wird ebenfalls in zwei Stufen ("zutreffend" und "andere Antworten als zutreffend") verrechnet. Abhängige Variablen sind die thematischen Fragen, aus denen die Antworten im PVF auspartialisiert werden.

Für die Wirkungen der Gesamtnoten ist keine bedeutsame Interaktion mit der erlebten Selbstverantwortlichkeit zu berichten. (Dies gilt auch für eine dimensionale Verrechnung der Selbstverantwortlichkeit.) Dagegen treten für die einzelfachbezogenen Fragen bei der (statistisch nicht unproblematischen) Auswertung "Über-alle-Fächer" mehrere signifikante Interaktionen auf. Um deren Richtung zu ermitteln, werden zusätzlich zur Kovarianzanalyse in den durch die Stufen des Faktors Selbstverantwortlichkeit gebildeten Gruppen von Prüfungen partielle Korrelationen zwischen Noten und thematischen Fragen in der üblichen Art berechnet.

Tabelle: III.III. - 10

Interaktion von Selbstverantwortlichkeit und Noten "Über-alle-Fächer"

	Wahrsch.zuf. Interaktion Note x PFF-F 10. WIRKL.LEISTUNG gegen PFF (PVF auspart.)	part.Korrelat. Note / PFF (PVF auspart.) PFF - F 10. WIRKL.LEISTUNG		Wahrsch.zuf. Interaktion Note x PFF-F 11. STUDIENLEISTUNG gegen PFF (PVF auspart.)	part.Korrelat. Note / PFF (PVF auspart.) PFF - F 11. STUDIENLEISTUNG	
		zutr. Note	nicht zutref.Note		zutr. Note	nicht zutref.Note
SK.FÄCHER						
Vordiplom		+.33	+.24	<.01 (?)	+.52	+.18
VD/nur psych.Fächer				<.05 (?)		
Hauptdiplom		+.18	+.23		+.09	+.27
MOT.W.A.						
Vordiplom		+.19	+.12		+.15	+.15
VD/nur psych.Fächer						
Hauptdiplom		+.20	+.08	<.01 (?)	+.33	-.04
MOT.BEL.						
Vordiplom		+.17	+.17		+.10	+.23
VD/nur psych.Fächer						
Hauptdiplom		+.32	+.17		+.33	+.17

Methodisches Vorgehen: Für jede der beiden Fragen sind drei Spalten vorgesehen. In der ersten Spalte werden die Ergebnisse des F-Tests der Interaktion in der zweifachen Kovarianzanalyse wiedergegeben. Ein "n.s." ist nicht vermerkt, da es wie die Angaben von Wahrscheinlichkeiten wegen der partiellen Abhängigkeit mit einem (?) versehen werden müßte. Die beiden folgenden Spalten führen die partiellen Korrelationen für die beiden durch den Faktor Selbstverantwortlichkeit gebildeten Gruppen von Prüfungen auf. Die Zahl der in der Kovarianzanalyse verrechneten Prüfungen beträgt im VD um 280, in den psychologischen Prüfungen des VD um 170 und im HD 145. Die partiellen Korrelationen beruhen jeweils auf unterschiedlichen Anzahlen von Prüfungen, sie dienen nur der Verdeutlichung der Richtung der Interaktion. Für die psychologischen Prüfungen des VD sind sie nicht gesondert berechnet worden. Positive Korrelationen zeigen eine Wirkung besserer Noten zu größerer Motivation und höherem Selbstkonzept an.

Die signifikanten Interaktionen betreffen mit dem Selbstkonzept eigener Fähigkeit im VD die thematische Frage, bei der auch in der Wirkungsanalyse die deutlichste Wirkung der Noten vorlag. Die Richtung der Interaktion entspricht den Erwartungen: Die Wirkung der Noten ist bei erlebter Selbstverantwortlichkeit größer. Auch die beim HD vorhandene Interaktion weist für die Handlungsplanung in diese Richtung, wenngleich wegen der geringen Notenstreuung diesem Ergebnis keine zu große Bedeutung beigemessen werden darf, weil es auf einige wenige Prüfungen zurückgehen kann. Die Moderatorwirkung ist an den Bezug der Selbstverantwortlichkeit zur erlebten Studienleistung gebunden. Dem ist noch hinzuzufügen, daß eine dimensionale Aufteilung der Selbstverantwortlichkeit (s. oben) eher weniger deutliche Ergebnisse erbringt. Das gilt sowohl für die Kovarianzanalyse als auch für die partiellen Korrela-

tionen.

Diese Moderatorwirkung ist wie schon bei der subjektiven Notenskala eine weitere Einschränkung der von den Normwerten geforderten universellen Wirkung der Noten.

Bei der zweifaktoriellen Kovarianzanalyse mit Selbstverantwortlichkeit und Noten als Faktoren wird ein zusätzlicher Haupteffekt der Selbstverantwortlichkeit deutlich: Die Selbstverantwortlichkeit hat eine von den Noten unabhängige "Wirkung" auf die thematischen Variablen.

Tabelle: III.III. - 11 (s.S. 316)

"Wirkung" der Selbstverantwortlichkeit
"Über-alle-Fächer"

Methodisches Vorgehen: Für die beiden Fragen sind jeweils drei Spalten vorgesehen. In der ersten findet sich das Ergebnis des F-Tests der zweifachen Kovarianzanalyse, bei der der Faktor Selbstverantwortlichkeit mit den Stufen "zutreffend" und "andere Antworten als zutreffend" vom Faktor Noten unabhängig geprüft wird (NIE et al. 1975). Abhängige Variablen sind wie üblich die thematischen Fragen, aus denen die Antwort im PVF auspartialisiert wird. Die partielle Korrelation in der zweiten der drei Spalten wird aus der Verrechnung der Selbstverantwortlichkeit als Dimension gewonnen; Note und Antwort im PVF werden auspartialisiert. Die Anzahl der Prüfungen ist identisch mit der Tabelle: III.III. - 10.

Positive Korrelationen bedeuten: Je mehr die Note in der Richtung "zu schlecht" vom Prüfling bewertet wird, um so größer ist die Änderung in Richtung größerer Motivation und höheren Selbstkonzepts.

Die im Ausmaß geringen, aber durchaus mit den

Tabelle: III.III. - 11

"Wirkung" der Selbstverantwortlichkeit "Über-alle-Fächer"

	Wahrsch.zuf. Zusammenhangs PFF - F 10. WIRKL.LEIST. gegen PFF (PVF auspart.)	part.Korr. PFF-F 10./ PFF (PVF und Noten auspart.)	Wahrsch. zuf.Abw. von $r_{part.} = .00$	Wahrsch.zuf. Zusammenhangs PFF - F 11. STUDIENLEIST. gegen PFF (PVF auspart.)	part.Korr. PFF-F 11./ PFF (PVF und Noten auspart.)	Wahrsch. zuf.Abw. von $r_{part.} = .00$
SK.FÄCHER						
Vordiplom	<.01 (?)	+.25	<.01(?)	<.01 (?)	+.22	<.01(?)
VD/nur psych.Fächer	<.05 (?)	+.23	<.01(?)	<.01 (?)	+.17	<.05(?)
Hauptdiplom	<.05 (?)	+.07		<.01 (?)	+.26	<.01(?)
MOT.W.A.						
Vordiplom		+.15	<.05(?)		+.10	
VD/nur psych.Fächer		+.11			+.14	
Hauptdiplom		+.11		<.05 (?)	+.23	<.01(?)
MOT.BEL.						
Vordiplom		+.06			-.02	
VD/nur psych.Fächer		+.10			+.06	
Hauptdiplom		+.05			+.11	

Wirkungen der Noten vergleichbaren Zusammenhänge der erlebten Selbstverantwortlichkeit mit den thematischen Fragen sind im wesentlichen auf das Selbstkonzept eigener Fähigkeit beschränkt. Sie sind Indikatoren für Änderungsprozesse, die auf das selbstbewußte Erleben der Bewältigung der einzelnen Prüfungen zurückgehen können. Dieses Erleben der erfolgreichen Bewältigung der Prüfung könnte die gemeinsame Grundlage der Änderungen des Selbstkonzepts und der Bewertung der in der Prüfung erzielten Note sein. (Die ähnliche Benennung eines Faktors in der Strukturanalyse aller Fragen entspricht dort nicht einem ähnlichen Zusammenhang von Selbstkonzept und Selbstverantwortlichkeit; s. Kapitel III.II.B.6.).

Bei der Auswertung des Zusammenhangs von erlebter Selbstverantwortlichkeit und Gesamtkonzepten, der von der Gesamtnote unabhängig ist, ist im VD in gleicher Richtung eine signifikante partielle Korrelation beim zukunftsbezogenen Selbstkonzept festzustellen. Das bestätigt die Ergebnisse für die Auswertung "Über-alle-Fächer"; es ist wie diese in seiner Besonderheit kaum ohne zusätzliche Datenerhebungen zu interpretieren. Auch der mögliche Unterschied von kategorialer Auswertung (Kovarianzanalyse) und dimensionaler (partielle Korrelation) bedürfte weiterer Nachforschungen.

Zusammenfassung (III.III.3.)

Die erlebte Selbstverantwortlichkeit für die Note ist in den Begriffsbereich der Ursachenzuschreibung einzuordnen. Sie wird als nicht personalistisch fixierter Begriff verwendet. Der personalistischen Verwendung entstammt die Erkenntnis, daß die Selbstverantwortlichkeit eine Moderatorfunktion bei der Verarbeitung von Rückmeldungen erfüllt.

Zur empirischen Erfassung werden fachbezogen zwei verschiedene Bezugssysteme als Grundlage der Selbstver-

antwortlichkeit vorgegeben: die erlebte Prüfung und das Studium. Die beiden Bezüge sind besonders im VD weitgehend voneinander abhängig. Die Antworten zu den einzelnen Fächern sind situationsspezifisch bestimmt, wenn auch für das VD eine mäßige personale Konstanz erkennbar ist. Die Verteilung der Antworten zeigt nur geringe Häufigkeiten für die "Unterschreitung" erlebten Zutreffens der Noten.

Die Moderatorfunktion der erlebten Selbstverantwortlichkeit ist bei der Auswertung "Über-alle-Fächer" unter einzelnen Bedingungen in der erwarteten Richtung deutlich: Beim Selbstkonzept im VD und bei der Handlungsplanung im HD ist die Änderung bei erlebter Selbstverantwortlichkeit größer. Dies ist nur der Fall, wenn die Note an der Einschätzung der eigenen Leistung im Studium gemessen wird. Zusätzlich dazu wird ein weiteres unerwartetes Ergebnis festgestellt: Die erlebte Selbstverantwortlichkeit weist einen von den Noten unabhängigen Zusammenhang mit Änderungen von Selbstkonzept und (geringfügiger) auch mit Änderungen von Handlungsplanung auf. Im VD ist ein solcher Zusammenhang im zukunftsbezogenen Gesamtkonzept gleichfalls vorhanden. Die Ergebnisse sind im einzelnen ursächlich nicht zu klären, entsprechen aber in ihrer Gesamtheit nicht den Normwerten.

III.III.4.: Die erlebte Rückmeldung in der Prüfung

Die erlebte Rückmeldung in der Prüfung wird nur fachbezogen erfragt; hier gelten die Fragen PFF - F 12. RÜCKM.NEG. und PFF - F 13. RÜCKM.POS. Sie stellt eine rein eklektische Fragestellung ohne begrifflichen oder empirischen Kontext dar. Die Verwendung der Frage im Zusammenhang der Stufenanalyse geht davon aus, daß hinter der Note, die vom Prüfer nach der mündlichen Prüfung erteilt wird, eine Prüfungsinteraktion steht, die vom Prüfling relativ unabhängig von der Note in

eigenen Bezugssystemen erlebt werden kann. Dieses Erleben kann gleichzeitig mit Änderungsprozessen bei den thematischen Variablen einhergehen. Es würde dann ähnlich der Selbstverantwortlichkeit eine Wirkung der Prüfung anzeigen, die nicht über die Noten vermittelt ist. (Dabei kann, wie in diesem Kapitel schon mehrfach betont, nicht angenommen werden, daß die erlebte Rückmeldung selbst die Ursache von Änderungen ist.)

Die erlebte Rückmeldung läßt sich in zwei Aspekte aufspalten, die positive und die negative Rückmeldung. Beide Aspekte kommen als Moderatorvariable und als alternative unabhängige Variable in Frage. Die Moderatorfunktion könnte darin bestehen, daß die Wirkung der Noten ähnlich wie an die erlebte Selbstverantwortlichkeit an die erlebte Begründung durch die Prüfungsinteraktion selbst gebunden ist. Die mögliche alternative "Wirkung" wird durch die Vermutung begründet, daß die Prüfungsinteraktion eine von den Noten unabhängige Information über die Leistung geben könnte.

Die statistische Auswertung erfolgt wie in Teil 3 über Kovarianzanalysen und partielle Korrelationen mit dem Unterschied, daß die Antwortskalen als quantitativ abgestufte Dimensionen zu behandeln sind. Sie werden über die Dichotomisierung am Median der Verteilung zu Stufen eines Faktors verwandelt.

Zur Vorbereitung dieser Auswertung sollen die Korrelationen zwischen den Fragen und innerhalb der Fragen Beachtung finden.

Tabelle: III.III. - 12

Korrelationen der Fragen zur erlebten Rückmeldung

	Vordiplom			Hauptdiplom		
	Kor-rel.	Streu-bereich von	bis	Kor-rel.	Streu-bereich von	bis
Korrelationen zwischen den Fragen						
PFF - F 12./PFF - F 13.						
"Fach-Prüfer-kombinationen" (mittlere Korrelation)	.24	-.21	.74	.18	-.07	.79
"Über-alle-Fächer"	.24			.17		
Korrelationen innerhalb der Fragen						
Paare von Einzelfächern (mittlere Korrelation)						
PFF - F 12.RÜCKM.NEG.	.22	-.25	.49	.11	-.15	.33
PFF - F 13.RÜCKM.POS.	.35	-.03	.53	.34	.12	.63

Methodisches Vorgehen: s. Tabelle: III.III. - 8.

Die Korrelation zwischen den beiden Fragen gibt Aufschluß darüber, ob die Antworten mehr im Sinne einer globalen Wertung der Prüfungsinteraktion erfolgen: Dann müßten deutliche negative Korrelationen zwischen ihnen bestehen; oder ob positive und negative Rückmeldung aneinander gebunden vorkommen und so eine informative Prüfung von einer wenig informativen Prüfung unterscheiden: Die beiden Antworten müßten dann deutlich positiv miteinander korrelieren.

Die empirisch ermittelten Werte der ersten beiden Zeilen zeigen, daß der positive und der negative Aspekt der Rückmeldung im Mittel weitgehend voneinander unabhängig sind. Rückmeldungen in Prüfungen sind also nicht eindimensional. (Die Noten stellen zur Mitteilung

aber nur ein eindimensionales System zur Verfügung.)

Zur Bestimmung, ob die Beantwortung der fachbezogenen Fragen mehr nach den einzelnen Prüfungssituationen differenziert oder personal konstant erfolgt, werden die mittleren Korrelationen zwischen allen Paaren von Einzelfächern berechnet (Zeile 3 und 4 der Tabelle: III.III. - 12). Beide Fragen werden eher situativ differenziert beantwortet, wenn auch die positive Rückmeldung Übereinstimmungen zwischen den Fächern aufweist, die gleichzeitig eine personal stabile Erlebnisweise der Prüfungsinteraktion als Varianzkomponente herausstellen.

Die Moderatorfunktion der erlebten Rückmeldung, die in der üblichen Kovarianzanalyse als Interaktion des Faktors Note mit dem Faktor Rückmeldung, der durch Dichotomisierung am Median gebildet wird, überprüft wird, ist nur in einer der 18 durchgeführten Kovarianzanalysen signifikant (die Zahl der Analysen ergibt sich aus den drei thematischen Fragen, die jeweils auf den drei Ebenen "Über-alle-Fächer": VD, VD/nur psychologische Fächer und HD berechnet werden). Eine Moderatorfunktion der erlebten Rückmeldung gilt als nicht nachgewiesen.

Dagegen sind für die alternative "Wirkung" einige im Ausmaß geringfügige, aber bei Verrechnung aller einzelnen Prüfungen signifikante Zusammenhänge zwischen erlebter Rückmeldung und thematischen Variablen vorhanden.

Tabelle: III.III. - 13

"Wirkung" der erlebten Rückmeldung "Über-alle-Fächer"

	Wahrsch.zuf. Zusammenhangs PFF - F 12. RÜCKM.NEG. gegen PFF (PVF auspart.)	part.Korr. PFF-F 12./ PFF (PVF und Noten auspart.)	Wahrsch. zuf.Abw. von $r_{part.} = .00$	Wahrsch.zuf. Zusammenhangs PFF - F 13. RÜCKM.POS. gegen PFF (PVF auspart.)	part.Korr. PFF-F 13./ PFF (PVF und Noten auspart.)	Wahrsch. zuf.Abw. von $r_{part.} = .00$
SK.FÄCHER						
Vordiplom		+.18	<.01(?)		+.01	
VD/nur psych.Fächer		+.16	<.05(?)	<.05 (?)	-.06	
Hauptdiplom		+.01			-.02	
MOT.W.A.						
Vordiplom		-.03			+.13	<.05(?)
VD/nur psych.Fächer		+.05		<.01 (?)	+.27	<.01(?)
Hauptdiplom		+.03			+.20	<.05(?)
MOT.BEL.						
Vordiplom		-.03			+.05	
VD/nur psych.Fächer		+.01			+.19	<.05(?)
Hauptdiplom		-.03		<.05 (?)	+.23	<.01(?)

Methodisches Vorgehen: Die Tabelle entspricht in Auswertung und Aufbau völlig der Tabelle: III.III. - 11.

Positive Korrelationen bedeuten bei zunehmender positiver und bei abnehmender negativer Rückmeldung mehr Änderung in der Richtung höheren Selbstkonzepts und größerer Motivation.

Trotz der Geringfügigkeit ist in den partiellen Korrelationen ein charakteristisches Muster zu erkennen: Das Ausbleiben erlebter negativer Rückmeldung ist mit einer eher positiven Änderung des Selbstkonzepts verbunden; dieser Zusammenhang beschränkt sich auf das VD. Das Ausmaß positiver Rückmeldung hat keine Beziehung zum Selbstkonzept, dafür aber sowohl im VD als auch im HD eine zur Motivation (Handlungsplanung und Beliebtheit). Im VD erreicht dieser Zusammenhang nur in den psychologischen Fächern ein etwas größeres Ausmaß. Dort war in der Wirkungsanalyse für die Noten keine Wirkung festzustellen. Die Ergebnisse zeigen, daß der Verlauf der Prüfung selbst in beschränktem Umfang Hinweise für die Handlungsplanung geben kann, wenn diese Hinweise erfolgsorientiert wahrgenommen werden können. Das entspricht jenen im theoretischen Teil berichteten Ergebnissen, daß Erfolg eher eine Wirkung auf die Motivation ausübt als Mißerfolg.

Eine abschließende Bemerkung zu den Ergebnissen der Stufenanalyse, sofern eine Auswertung auf der Ebene "Über-alle-Fächer" vorgenommen wird, die die einzelnen Prüfungen und damit jeden Prüfling mehrmals in die Auswertung mit einbezieht, ist angebracht: Zwar kann in den vorbereitenden Auswertungsschritten jeweils nachgewiesen werden, daß die Antworten vorwiegend situationsspezifisch gegeben werden. Doch sind die meist geringfügig positiven Korrelationen zwischen den Antworten zu den einzelnen Fächern ausreichend, um bestimmten personalen Tendenzen einen Einfluß auf die Ergeb-

nisse zu ermöglichen. Dadurch aber können einige wenige Personen die Ergebnisse beeinflussen, was dann zu falschen Einschätzungen in der inferenzstatistischen Betrachtung führen muß. In der Wirkungsanalyse konnte vor solchen Fehlschlüssen mittelbar der Vergleich mit den Ergebnissen auf der Ebene von Gesamtwerten, bei denen jeder Prüfling nur einmal in die Auswertung einbezogen wird, zumindest für das Selbstkonzept bewahren. Ferner war die Zuordnung zu theoretischen Erwartungen eine weitere Stütze. Bei den Teilen 2, 3 und 4 dieses Kapitels sind solche Sicherungen zwar auch noch an einigen Stellen, aber spärlicher vorhanden: Die Ergebnisse dürfen deshalb nicht als mehr als erste Hinweise betrachtet werden.

Zusammenfassung (III.III.4.)

Die erlebte Rückmeldung bezieht sich auf die Interaktion in den mündlichen Prüfungen; sie wird in einen Aspekt positiver und einen Aspekt negativer Rückmeldung aufgespalten. Es wird vermutet, daß sie als Moderatorvariable und als alternative unabhängige Variable wirksam werden kann.

Die Auswertung zeigt, daß die beiden Aspekte weitgehend voneinander unabhängig sind. Die negativ erlebte Rückmeldung in der Prüfung ist im wesentlichen situationsspezifisch determiniert; bei der positiv erlebten Rückmeldung ist eine personal stabile Tendenz deutlicher ausgeprägt.

Eine Moderatorfunktion in Interaktion mit den erteilten Noten ist für die erlebte Rückmeldung nicht nachzuweisen. Dagegen ist eine im Ausmaß zwar geringe, aber statistisch bei der Auswertung "Über-alle-Fächer" bedeutsame "Wirkung", die von den Noten unabhängig ist, unter verschiedenen Bedingungen festzustellen. Das Ausbleiben negativer Rückmeldung ist im VD mit einer posi-

tiven Änderung des Selbstkonzepts verbunden. Die Zunahme positiv erlebter Rückmeldung in der Prüfung hat einen positiven Einfluß auf die Motivation. Eine abschließende Bemerkung betont, daß in der Auswertung "Über-alle-Fächer" in 3 Teilen dieses Kapitels mit fragwürdigen Mitteln Ergebnisse gewonnen werden, die nicht mehr als erste Hinweise sein können.

IV.: Praktische Konsequenzen für die Rückmeldung im Studium

Die Formulierung praktischer Konsequenzen ist vor die Entscheidung gestellt, diese mehr auf die Prüfung als institutionellen Rahmen, der den Ausgangspunkt der Arbeit darstellt, oder mehr auf die Funktion der Rückmeldung, die sich im Laufe der einbettenden theoretischen Analysen in Kapitel II immer weiter verselbständigt hat, zu beziehen. Die Verselbständigung ging einher mit der Einsicht, daß die Prüfung dominant anderen Funktionen wie der Selektion und der Sozialisation zugeordnet ist. Die Zuordnung, so lautete der Schluß, macht eine auf die Rückmeldung an den Prüfling zentrierte Reform der Prüfung weder möglich noch durchsetzbar. Das aber würde bedeuten, daß die spezifischen Ergebnisse der Untersuchung in praktische Konsequenzen nicht einzubringen wären, wenn diese im Hinblick auf die Prüfung formuliert werden. Deshalb beschränken sich die Überlegungen zu praktischen Konsequenzen auf die Rückmeldung des Lernerfolgs an die Studenten. Dabei steht das Selbstkonzept eigener Leistungsfähigkeit im Vordergrund.

Die Notwendigkeit praktischer Änderungen ist daraus abzuleiten, daß die Rückmeldungen in den vorliegenden Prüfungen weder sinnvoll noch genügend wirksam erfolgen. Die Wirksamkeit der Rückmeldung wurde an Normwerten überprüft. Dabei stellte sich heraus, daß sich Motivation und Selbstkonzept zwar in der Rich-

tung ändern, die die Normwerte verlangen. Bis auf das Selbstkonzept eigener Fähigkeit im VD bleiben diese Wirkungen im Ausmaß äußerst geringfügig. Aber auch beim Selbstkonzept ist das Ausmaß nicht so groß, daß die Normwerte als erfüllt gelten können. Zwar werden die empirischen Ergebnisse teilweise mit fragwürdigen statistischen Methoden gewonnen, die den Daten nicht voll angemessen sind. Im HD ist ihre Aussagekraft noch weiter durch die deutliche Selektion der Beantworter eingeschränkt. Doch trägt das Ergebnis der theoretischen Analyse zur Sicherheit des abschließenden Urteils bei, weil es zu Erwartungen Anlaß gibt, die den empirischen Ergebnissen ungefähr entsprechen.

Die Feststellung des Sinns der Wirkungen setzt vor den funktionalen Verknüpfungen bei den Schritten an, die zwischen der Formulierung der Normwerte und der Wirkungsanalyse liegen. Es wurde herausgearbeitet, daß unter Berücksichtigung der Kritik am Leistungsprinzip für die vergleichsorientierte Kennzeichnung der Leistungen, die Noten liefern, einige einschränkende Bedingungen zu fordern sind, die erst eine positiv zu wertende Verwendung dieser Noten erlauben. Diese Bedingungen aber werden durch die bestehenden Prüfungen und Noten nur unzureichend erfüllt. Ein weiterer Schritt im Rahmen der Voraussetzungsanalyse ergab an Hand der Auswertung fremder Untersuchungen und des hier vorliegenden Notenmaterials, daß eine Reihe von Eigenschaften die Noten als Grundlage der Rückmeldung wenig geeignet erscheinen läßt.

Angesichts der Gefahren der Mißinformation durch Noten, der geringen inhaltlichen Validität von Prüfungen und der sozialen Problematik des unzureichend eingegrenzten Prinzips konkurrenzhafter Leistungsvergleiche könnte man die bisher berichteten Ergebnisse der Untersuchung mit der Zusammenfassung abschließen: Die für die Prüfungen nicht zentrale Funktion der Rück-

meldungen an die Prüflinge wird in Form fragwürdiger Konkretisierung (der Normwerte) mit fragwürdigen Mitteln (den Noten) nur in so geringem Ausmaß erfüllt, daß dadurch wenig Schaden angerichtet wird.

Eine vollständigere Aufarbeitung der Ergebnisse aber läßt diesen beruhigenden Schluß nicht zu. Bei der Wirkung der Noten auf das Selbstkonzept im VD ist darauf zu verweisen, daß in einzelnen "Fach-Prüfer-Kombinationen" auch im Ausmaß bedeutsame Wirkungen auf das Selbstkonzept vorliegen. Ein Ergebnis der Stufenanalyse ist, daß bei erlebter Selbstverantwortlichkeit für die Note die Wirkung im Ausmaß gleichfalls erheblich ist. Dabei erfolgen die Wirkungen unter Bedingungen, die einer solchen Wirkung eher abträglich sein sollten, wie der geringen Differenzierung der Notenverteilungen und der negativen Einstellungen der Prüflinge gegen die Prüfungen (TIMMERBRINK 1971). Das läßt den Schluß zu, daß das Selbstkonzept, besonders im VD, vom Studium her so wenig stabilisiert ist, daß es auch unter diesen nicht optimalen Bedingungen den Einflüssen der Noten ausgesetzt ist.

Dieser Eindruck verstärkt sich, wenn nicht allein die Noten als Quellen der Wirkung in Betracht gezogen werden. Die Wirkungsanalyse ergibt, daß sich infolge der Prüfung die fachbezogenen Selbstkonzepte entdifferenzieren; sie ergibt ferner, daß der sinnvolle Zusammenhang von Selbstkonzepten und Handlungsplanung nach der Prüfung eher geringer ist als vor der Prüfung. Die Stufenanalyse macht durch den Nachweis von Moderatorwirkungen der subjektiven Notenskalen und der erlebten Selbstverantwortlichkeit auf Verfälschungen der Einflüsse aufmerksam. Die berichteten Zusammenhänge der Selbstkonzepte mit der erlebten Rückmeldung und der erlebten Selbstverantwortlichkeit sind als weitere Anzeichen der Empfänglichkeit der Selbstkonzepte für Einflüsse zu deuten. Nicht zuletzt beweist die Faktoren-

analyse die fehlende Einbettung des Selbstkonzepts in übergreifende Erlebnisstrukturen. Diese Beobachtungen sind vorwiegend auf das VD beschränkt. Das kann auf die in der Voraussetzungsanalyse vermutete Verankerung des Selbstkonzepts des HD in Kompetenzerfahrungen des Studiums zurückgeführt werden, wenngleich auch im HD einige dieser unerwünschten Wirkungen von Noten und Prüfungen auf das Selbstkonzept vorliegen.

Die Behauptung von der fehlenden Stabilisierung des Selbstkonzepts eigener Fähigkeit im Studium wird durch die Ergebnisse einer Befragung von Lehrenden bestätigt, die in ihrer Mehrheit bezweifeln, daß die Studenten ihre Fähigkeiten realistisch einschätzen (CLEMENS-LODDE 1974). Der ungenügenden Realistik und Stabilität des Selbstkonzepts steht die Einsicht gegenüber, daß ein differenziertes, realistisches und stabiles Selbstkonzept, das sich bei wenig legitimen Einflüssen nicht ändert, eine zentrale Bedeutung für das selbstgesteuerte Lernen hat. Im theoretischen Teil (Kapitel III.I.C.) wurden Untersuchungen zur Bedeutung des Selbstkonzepts für erwartungswidrig schlechte Leistungen (underachievement) und für die Planung des eigenen Lernverhaltens angesprochen. Umfassender, aber nicht in Einzelheiten empirisch abgesichert, stellt die Selbstkonzept-Theorie (schulischen) Lernens die Bedeutung des Selbstkonzepts heraus (WILHELMS 1967). Sie nimmt an, daß das Selbstkonzept Ausgangspunkt und Ziel selbstverwirklichenden Lernens ist. Auf der Grundlage bestehender Selbstkonzepte werden Lernziele gesetzt und verwirklicht; die Rückmeldungen über die Ergebnisse der Verwirklichung werden in das Selbstkonzept integriert. Dieser Entwurf entspricht dem Rahmenziel selbstbestimmten Lernens; die geforderte aktive Passung von eigenen Fähigkeiten und Anforderungen ist als notwendiger Bestandteil eines derartig gesteuerten Lernprozesses zu sehen. Dem Entwurf folgend ist eine

Neubestimmung von Struktur, Gegenständen und Inhalten des Selbstkonzepts notwendig. Es ist inhaltlich auszudifferenzieren, indem an die Stelle der Fächer sachlich differenzierte Strukturbildungen treten, die auch praktische Aspekte von Tätigkeiten umfassen können. Der Aspekt vergleichender Leistungsgüte müßte durch Bezugssysteme sachlich angemessener Problemlösungen ersetzt werden.

Das so verstandene Selbstkonzept hat nicht nur eine reflexive, sondern auch eine dynamische Bedeutung, indem es die Grundlage von Zielsetzungen darstellt. Insofern wird von den Überlegungen zum Selbstkonzept auch die Motivation, insbesondere als Handlungsplanung, tangiert. In Kapitel III. II.B.4. wurde mit dem Hinweis auf den sinnvollen Zusammenhang von Selbstkonzept und Handlungsplanung diese Integration schon angedeutet. Die Integration geht über den Versuch einer leistungsperspektivischen Verbindung von Selbstkonzept und Motivation (MEYER 1973) hinaus, weil die Ziele inhaltlich als Interessen und nicht nur formal durch ihre Leistungshöhe zu qualifizieren sind. Rückmeldungen sollen im Rahmen des Entwurfs zur Differenzierung und Realistik des Selbstkonzepts beitragen. In dieser Funktion müssen die Rückmeldungen, weil sie auf selbstbestimmte Steuerungsprozesse eigenen Lernens folgen, sich auf die Ziele beziehen, die diesen Steuerungen zugrunde liegen (KVALE 1972; SCHREINER 1972; WINGO 1960).

Der konstruktive Beitrag der Psychologie kann wiederum methodischer Art sein: Sie kann zur Entwicklung von Methoden der Rückmeldung beitragen, die Realistik und Differenzierung des Selbstkonzepts gewährleisten. Ein Ansatzpunkt in diesem Sinne wirksamer Rückmeldung ist die Entwicklung einer wirkungsvollen Beratung und Anleitung (guidance) der Lernenden (KANDEL 1936), eine Fragestellung, die auch Konzepten pädagogischer Diagnostik eng verbunden ist (HOPF 1975).

Arbeiten zur Verwendung von Tests im pädagogischen Bereich betonen die Notwendigkeit der Individualisierung und des konkreten Zielbezugs bei der Beratung auf der Grundlage von Testergebnissen (DAVIS 1971; LISTER & McKENZIE 1968; THORNDIKE & HAGEN 1969). Der Übersicht von TYLER (1969) ist zu entnehmen, daß allein im Bereich von Kurzberatungen für Schüler und Studenten systematische Erfahrungen mit der Auswirkung der Beratung auf das Selbstkonzept sich angesammelt haben. Diese Methode der (Bildungs-)Beratung (häufig unter dem Begriff des "brief educational-vocational counseling" zusammengefaßt) entspricht nicht voll dem sinnvollen Stellenwert von Rückmeldungen in dem Entwurf selbstkonzeptorientierten Lernens, da ihr Gegenstand nicht auf konkrete Lernziele in der Form individueller Ziele abgestimmt ist, sondern zu einem bestimmten Zeitpunkt (meist zu Beginn der Aufnahme eines Studiums in der Orientierungsphase) universell erfolgt. In methodischer Hinsicht aber sind Erfahrungen damit durchaus wichtige Hinweise.

Eine zeitlich nicht sehr umfangreiche Kurzberatung auf der Grundlage von Testergebnissen kann spezifische Selbstkonzepte (wie auch das globale Selbstkonzept, was hier weniger interessiert) in der Richtung größerer Realistik beeinflussen (FROEHLICH 1957; JOHNSON 1953; LISTER & OHLSON 1965; TORRANCE 1954; WILLIAMS 1962). Wenige Untersuchungen berichten allerdings auch Mißerfolge solcher Versuche (BERDIE 1954, der aber indirekte Auswirkungen an der Persistenz feststellt; FOLDS & GAZDA 1966; ROBERTSON 1960). Die Zahl der Untersuchungen mit positiven Ergebnissen aber ist noch um die folgend aufgeführten zu vermehren: Unterschiedliche Beratungsmethoden haben kaum differenzierte Wirkungen (HOLMES 1964); wie auch die Effekte von Gruppenberatungen sich kaum von Einzelberatungen unterscheiden (CATRON 1966; WALKER 1965; WRIGHT 1963).

Einzelberatungen werden allerdings eindeutig von den Studenten bevorzugt (FOLDS & GAZDA 1966; HOLMES 1964; WALKER 1965; WRIGHT 1963).

Wenn die quantitativen Effekte der Beratungen betrachtet werden, sind die Änderungen jedoch im Ausmaß deutlich begrenzt. Als methodische Alternative zu der einfachen Mitteilung von Ergebnissen, wie sie bei Prüfungen üblich ist, scheinen sie diesen aber überlegen. HILLS & WILLIAMS (1965) können absichern, daß die einfache Mitteilung von Testergebnissen ohne Beratung keine entsprechenden Wirkungen hat. Mit Bezug auf die begrenzte Wirkung ist auch zu betonen, daß meist ein strenges Kriterium des Beratungserfolgs (Zunahme der Realistik als absolute Übereinstimmung mit Testergebnissen) verwendet wird, das strenger ist als das in dieser Untersuchung für die Noten verwendete der relativen Übereinstimmung. Die positive Wirkung der Beratung gegenüber der einfachen Mitteilung ist u.a. darin begründet, daß über die Kommunikation auch die Verarbeitung der Bewertung beeinflußt werden kann (DAVIS 1971). Beratungen können in einem besonderen Bereich wirksam werden, der Noten und Prüfungen sicher verschlossen bleibt: der Ermutigung von Studenten mit geringen Fähigkeiten. Nach ROSS (1938) und MEICHENBAUM & SMART (1971) können ermutigende Beratungen, die die gegenwärtigen Schwächen nicht verschweigen, den Studienerfolg von Studenten mit schlechten Prognosen deutlich erhöhen.

Die Ausrichtung an dem Entwurf der Selbstkonzept-Theorie des Lernens macht die Verlagerung der Beratung in das Studium entsprechend dem Plan des "diagnostic teaching", der aus dem selbstkonzeptorientierten Lernens abgeleitet wird (WILHELMS 1967), sinnvoll. Eine Integration des Konzepts des "Erfahrungslernens" (TORBERT 1973) scheint auf weite Sicht möglich.

Eine vorrangigere Überlegung muß den möglichen

Quellen der Informationen für solche Beratungen gelten. Die gegenwärtigen (selektiven) Prüfungen sind wegen ihrer dargestellten Nachteile als Grundlage wenig brauchbar. Als Möglichkeiten der Informationsgewinnung bieten sich die lernzielorientierte Leistungsmessung (STRITTMATTER 1973) und informierende Prüfungen (KAPUSTE 1970) an. Einen wichtigen Stellenwert hat auch die Sammlung von Informationen über den Studienverlauf (HARDEE 1960), die an deutschen Universitäten weitgehend fehlt, wenngleich die Gefahr nicht zu übersehen ist, daß solche Sammlungen von Daten zu leicht eher den Zwecken der Verwaltung (Reglementierung des Studiums) als der Förderung sachorientierten Lernens dienen (WILHELMS 1967).

Die Zuordnung der Beratung zum Entwurf selbstkonzeptorientierten Lernens macht eine von der Lehre abgelöste Institutionalisierung der Beratung, wie sie in Anlehnung an amerikanische Vorbilder versucht wird, wenig sinnvoll. Die Beratung wird als notwendiger Teil der Organisation des individuellen Lernprozesses verstanden und muß zugleich ein integrierter Bestandteil der Lehre sein. Nicht zuletzt ist das auch für die Lehrenden von Vorteil, da der Beratungsprozeß nicht auf einen einseitigen Informationsfluß beschränkt bleiben muß, sondern dem Lehrenden Informationen über die Qualität seiner Lehre vermitteln kann.

Die Verwirklichung dieses Entwurfs macht aber auch studienorganisatorische Konsequenzen notwendig. Rückmeldungen über die selbstgesetzten Lernziele und damit zusammenhängend das Lernen aktiver Passung von Fähigkeiten und Anforderungen sind nur dort möglich, wo das Studium Freiheit für die Planung eigenen Lernens (und durch alternative Angebote gewisse Hilfen dazu) bietet. Da das Selbstkonzept gleichzeitig Grundlage der Berufswahl werden kann und für Lernprozesse im Rahmen praktischer Tätigkeit Bedeutung haben kann,

ist eine Integration von praxisbezogenen Lernangeboten sinnvoll (BAHRDT 1972). Gerade im Zusammenhang mit praxisbezogenem Lernen ist auch zu prüfen, wie weit die Rückmeldung über Beratungen durch direkte, sachorientierte Rückmeldung ergänzt oder abgelöst werden kann.

Die Entwicklung eines realistischen und sachlich differenzierten Selbstkonzepts im Studium könnte gerade im Hinblick auf die Prüfungen einen wichtigen Nebeneffekt haben: die Immunisierung des Selbstkonzepts gegen die abträglichen Wirkungen der Noten und der Prüfungen. Eine durch inhaltliche Differenziertheit und viele einzelne Erfahrungen gebildete komplexe Struktur von Selbstkonzepten müßte die anfänglich dargestellten negativ bewerteten Wirkungen der Noten und Prüfungen auf das Selbstkonzept mindern. Diese Erwartung aber gründet auf der Annahme, daß das Studium einen eigenständigen Wert erlangen kann, der von der Berechtigungsfunktion des Studiums (dem Erwerb eines meist über eine Berufsposition vermittelten gehobenen gesellschaftlichen Status) sich ablösen kann. Auf der Ebene des Selbstkonzepts entspricht dem die Erwartung, daß die Entwicklung eines sachangemessenen, realistischen Selbstkonzepts auch dann möglich und sinnvoll ist, wenn ein unrealistisch überhöhtes und eher undifferenziertes Selbstkonzept auf dem Markt der Berufschancen die bessere Ausgangsposition einräumt.

Zusammenfassung (IV.)

Die Formulierung praktischer Konsequenzen erfolgt im Hinblick auf die Funktion der Rückmeldung für das Selbstkonzept. Sie geht davon aus, daß die Ergebnisse der theoretischen und empirischen Analysen in dem Urteil zusammengefaßt werden können, daß die Rückmeldung durch die Prüfung weder sinnvoll noch wirksam

erfolgt. Es sind Wirkungen der Noten (und Prüfungen) festzustellen, die die Herausbildung eines realistischen und differenzierten Selbstkonzepts verfehlen. Dem steht mit der Selbstkonzept-Theorie des Lernens ein Entwurf gegenüber, der von der grundlegenden Bedeutung von realistischen und differenzierten Selbstkonzeptstrukturen für selbstverwirklichendes Lernen ausgeht. Als Beitrag zur Herausbildung solcher Selbstkonzepte werden zielbezogene Beratungen im Studium vorgeschlagen. Beratungen haben sich als wirksame Mittel zur Entwicklung realistischer Selbstkonzepte erwiesen. Die Beratungen sollten in das Studium integriert sein. Sie werden nur im Zusammenhang mit studienorganisatorischen Änderungen ihre Funktion im selbstkonzeptorientierten Lernen erfüllen können (dazu gehört auch die Integration von Praxiserfahrungen). Von einem stabilisierten Selbstkonzept wird eine Immunisierung gegen die abträglichen Wirkungen der Noten und Prüfungen erwartet. Das ist aber davon abhängig, daß sachliche Ansprüche des Studiums sich von seiner Funktion als Mittel des Statuserwerbs befreien können.

V.: Zusammenfassung der Arbeit

Dieses Kapitel hat sich nicht die Aufgabe gestellt, die Einzelheiten der theoretischen und empirischen Analyse darzustellen; dem Zweck dienen die Zusammenfassungen, die die einzelnen Teile der Arbeit abschließen. Auf sie wird durch Angabe der Seitenzahlen verwiesen.

Kapitel I.: Einführung in die Themenstellung

Die Arbeit untersucht mit den theoretischen und empirischen Mitteln der Psychologie die Wirkung des Vollzugs der akademischen Prüfung auf Selbstbewertung und Motivation der Prüflinge. Sie ist im Rahmen eines von der Stiftung Volkswagenwerk geförderten Forschungsprojekts am Psychologischen Institut der Universität Münster entstanden.

Kapitel II.: Hochschulpädagogische Vorüberlegungen und Hintergründe

(Zusammenfassungen: S. 16, 39, 48, 59)

Oberstes Ziel des methodisch orientierten, auf Reform ausgerichteten Ansatzes im Rahmen der Hochschuldidaktik ist die Ablösung von Fremdbestimmung durch Selbstbestimmung. Dem trägt besonders die Forderung Rechnung, daß die Rückmeldung in der Prüfung zur Entwicklung einer realistischen Selbstbewertung beitragen soll. Ferner wird gefordert, daß die Prüfung bei zu geringen und besonders guten Leistungen die Motivation erhöhen soll. Beides soll zu einer besseren Passung von individuellen Fähigkeiten und Anforderungen führen. Eine einbettende Funktionentheorie macht deutlich, daß die Rückmeldung an den Prüfling nur eine Nebenfunktion der Prüfung darstellt, deren Hauptfunktionen auf der

soziologischen Betrachtungsebene als Selektion und Sozialisation bestimmt werden können; auf der psychologischen Ebene der Betrachtung stellen sich diese als Steuerungs- und Diagnosefunktion dar. Die Funktionsanalyse ergibt, daß die Prüfung als Funktionenmischung bei Überformung durch die Selektionsfunktion kaum die Möglichkeit einer isolierten Reform zur Optimierung der Rückmeldung eröffnet. Die Noten als institutionelles Mittel der Rückmeldung liefern begrenzt Informationen (über die vergleichende Leistungsgüte), die nur unter bestimmten Bedingungen, die von den Prüfungen nicht genügend erfüllt werden, eine gerechtfertigte Grundlage der Rückmeldung sind.

Kapitel III.: Psychologische Analyse der Rückmeldefunktion

Die psychologische Analyse der Rückmeldefunktion geht in drei Schritten vor. In der Vorbereitungsanalyse werden die abhängigen Variablen Selbstbewertung als Selbstkonzept eigener Leistungsfähigkeit und Motivation als Handlungsplanung und Beliebtheit sowie die unabhängigen Variablen als Noten (und Prüfungen) theoretisch und empirisch näher bestimmt. In der Wirkungsanalyse werden für die globale Wirkung der Noten und der Prüfung auf Selbstkonzept und Motivation das Ausmaß und die Rahmenbedingungen dieser Wirkung untersucht. In der Stufenanalyse werden vermittelnde Prozesse in die Auswertung des empirischen Materials einbezogen.

Neben Erkenntnissen, die aus verschiedenen Gebieten der Psychologie gewonnen werden, dient der Analyse eine eigene empirische Untersuchung. Sie wird als Befragung vor und nach den mündlichen Prüfungen mittels eines Fragebogens bei Prüflingen am Psychologischen Institut der Universität Münster durchgeführt. Es wer-

den Prüflinge des Vordiploms (folgend VD) und des Hauptdiploms (HD) befragt; im HD liegt ein Selektionseffekt vor, da nur Prüflinge mit besseren Leistungen den Fragebogen beantworten; das schränkt die Verallgemeinerbarkeit der Ergebnisse des HD ein.

Kapitel III.I.: Voraussetzungsanalyse

(Zusammenfassungen: 78, 96, 106, 116, 140, 153, 160, 173, 184)

Die Noten weisen eine Reihe von Eigenschaften auf, die sie als Grundlage der Rückmeldung fragwürdig erscheinen lassen: Die Notenstufen sind nicht gleichabständig; die Verteilungen sind bedingungsabhängig; Zuverlässigkeit und Validität der Noten sind (besonders bei mündlichen Prüfungen) unzureichend; die Bildung von Gesamtnoten durch Mittelung ist problematisch.

Die empirische Auswertung der vorliegenden Noten stellt eine Verteilungsform seltener Ereignisse fest. Das ist Anlaß, für die Motivation die Forderung einer zusätzlichen kompensatorischen Wirkung aufzugeben. Die Noten in den einzelnen "Fach-Prüfer-Kombinationen" (den Prüfungen, die ein Prüfer in einem Fach abnimmt) sind sowohl im VD als auch im HD unterschiedlich verteilt.

Die Selbstbewertung wird theoretisch als Selbstkonzept eigener Leistungsfähigkeit bestimmt. Sie wird sowohl für die einzelnen Fächer als auch für die Psychologie insgesamt (gegenwärtiges/zukunftsbezogenes Gesamtkonzept) erfaßt. Als Methode der Abfrage ist ein Vergleich mit den Mitprüflingen vorgesehen. Die vorliegende Literatur läßt die Validität des Konstrukts im Sinne der Zielvorstellungen als gesichert erscheinen, wenngleich auch die Möglichkeit systematischer Fehler besteht.

Die Motivation wird in die beiden Aspekte der Beliebtheit und der Handlungsplanung aufgespalten, letztere weist den direkteren Bezug zu den Zielvorstellungen auf. Beliebtheit und Handlungsplanung werden nur für die einzelnen Fächer erfragt. Eine Absicherung von Inhalt und Methode gelingt nicht so gut wie beim Selbstkonzept.

Die vorbereitende Auswertung der Daten beschränkt sich fast ausschließlich auf die Befragung vor den Prüfungen. Sie zeigt, daß die Antworten zum Selbstkonzept nicht verteilungsorientiert, sondern als globale Bewertungen abgegeben werden, die im Mittel Überschätzungen darstellen. Besonders im HD scheinen die Antworten auf Kompetenzerfahrungen aus dem Studium zurückzugehen. Die fachbezogenen Selbstkonzepte stimmen deutlich personal gebunden über die Fächer hinweg miteinander überein, während die Antworten zur Motivation im wesentlichen situativ verankert sind.

Kapitel III.II.: Wirkungsanalyse

(Zusammenfassungen: S. 208, 223, 243, 250, 255, 263, 267, 278)

Die Aufarbeitung von Ergebnissen der Forschung zur Verarbeitung von Fremdbewertung, zur Wirkung diagnostischer Informationen und einzelner Arbeiten über Noten macht eine begrenzte Wirkung der Noten auf das Selbstkonzept in der Richtung der Normwerte wahrscheinlich. Gleichzeitig aber lassen die Erkenntnisse erwarten, daß bestimmte Situations- und Persönlichkeitsbedingungen zu Abweichungen von der normentsprechenden Wirkung beitragen.

Für die Motivation liegen noch zwingendere Hinweise auf im Sinne der Normwerte verfälschend wirkende Persönlichkeitsbedingungen vor. Entsprechend ist eine Wirkung im Sinne der Normwerte nicht so sicher wie

beim Selbstkonzept.

Die empirische Auswertung findet unter den verschiedenen Auswertungsbedingungen Wirkungen, die in die durch die Normwerte vorgegebene Richtung weisen. Aber nur bei den Gesamtkonzepten und fachbezogenen Selbstkonzepten des VD sowie dem zukunftsbezogenen Gesamtkonzept des HD ist eine Wirkung von etwas bedeutsamerem Ausmaß vorhanden. Alle Wirkungen sind quantitativ eher geringer, als es normativ gesehen sinnvoll ist (obwohl eine klare Quantifizierung der Normwerte zur Stützung dieses Urteils nicht möglich war).

Anschließende Auswertungsschritte zeigen, daß die Wirkung nicht wesentlich von dem Geschlecht der Prüflinge und im VD von der gewählten Prüfungsordnung abhängt. Dagegen ist das Ausmaß der Wirkung auf das Selbstkonzept im VD von der "Liberalität" der Prüfer bestimmt: Bei liberalerem Prüfungsstil nimmt das Ausmaß der Wirkung zu.

Als von den Noten unabhängige Wirkungen der Prüfungen werden ein Auseinanderfallen von Selbstkonzept und Motivation nach der Prüfung und eine Entdifferenzierung der fachbezogenen Selbstkonzepte gefunden. Beides ist kaum als positiver Beitrag der Prüfung zu bewerten. Diese Bewertung gilt auch für die Erhöhung der Mittelwerte der Selbstkonzepte, die schon vor der Prüfung Überschätzungen darstellen.

Eine abschließende Faktorenanalyse der fachbezogenen Antworten trägt zur Aufklärung möglicher Hintergründe der Ergebnisse bei. Sie zeigt, daß die motivationalen Fach-Bezüge über die Prüfung hinweg konstant bleiben. Die Note wird eher über eine wertende Stellungnahme zur Prüfung verarbeitet. Im VD ist das Selbstkonzept weitgehend isoliert, im HD dagegen zeichnet es sich durch eine über die Prüfung hinweg stabile Komponente aus, die mit der Leistungsbewertung in der

Prüfung übereinstimmt.

Kapitel III.III.: Stufenanalyse

(Zusammenfassungen: S. 297, 306, 317, 324)

Die Stufenanalyse beschränkt sich auf die isolierte empirische Prüfung einzelner vermittelnder Prozesse. Die subjektiven Notenskalen werden über eine Nutzenskalierung gemessen. Sie interagieren beim gegenwärtigen Gesamtkonzept im VD als Moderatorvariablen mit den Einflüssen der Gesamtnoten. Im HD ist eine solche Wirkung nicht nachzuweisen, statt dessen aber eine Änderung der subjektiven Notenskalen in Abhängigkeit von den Gesamtnoten. Besonders die Moderatorwirkung widerspricht als Einschränkung einer universellen Wirkung den Normwerten. Eine deskriptive Auswertung der subjektiven Skalen macht auf die dominante Mißerfolgsorientierung der Prüflinge aufmerksam und begründet so die geringe Wirkung unterschiedlich "guter" Noten.

Aus dem Anspruchsniveau, das vor der Prüfung als konkrete Erwartung erfragt wird, wird als Differenz zu den wirklichen Noten ein Maß für den subjektiven Erfolg und Mißerfolg gewonnen. Dieser subjektive Erfolg/Mißerfolg zeigt weniger Wirkung auf Selbstkonzept und Motivation als die wirklichen Noten und ist deshalb als wesentliche verfälschende Komponente subjektiver Verarbeitung der Noten nicht in Betracht zu ziehen.

Die erlebte Selbstverantwortlichkeit für die Note erweist sich im VD als Moderatorvariable bei der Wirkung der Noten auf das Selbstkonzept und im HD bei der Wirkung der Noten auf die Handlungsplanung. Die Wirkung ist jeweils bedeutsamer, wenn die Noten, an der Einschätzung der eigenen Leistung im Studium gemessen, als gerechtfertigt erlebt werden. Zusätzlich findet sich ein nicht erwarteter direkter Zusammenhang von

erlebter Selbstverantwortlichkeit und Änderungen des Selbstkonzepts. Dieser Zusammenhang ist von den erteilten Noten unabhängig.

Die erlebte Rückmeldung in der Prüfung wird in positiver und negativer Form erfragt. Eine Moderatorwirkung liegt nicht vor. Unabhängig von den Noten geht das Ausbleiben negativer Rückmeldung einher mit größerer Zunahme des Selbstkonzepts, während die Motivation eher mit dem Ausmaß positiver Rückmeldung steigt.

Das Ausmaß der Wirkungen und Zusammenhänge in der Stufenanalyse ist insgesamt gering; die statistischen Auswertungen erfolgen häufig mit fragwürdigen Methoden. Das letztere gilt teilweise auch für die Wirkungsanalyse. Die Ergebnisse der empirischen Untersuchung können deshalb nicht mehr als erste Hinweise darstellen.

Kapitel IV.: Praktische Konsequenzen
für die Rückmeldung im Studium

(Zusammenfassung: S. 333)

Die Erörterung von praktischen Konsequenzen geht von der zentralen Bedeutung eines realistischen und differenzierten Selbstkonzepts für das Lernen aus. Besonders im VD lassen die Ergebnisse den Schluß zu, daß die Prüfungen zu wenig und verfälschend zu einem realistischen Selbstkonzept beitragen, es auf der anderen Seite aber auch im bisherigen Studium nicht genügend stabilisiert worden ist. Eine diesbezügliche Reform der Prüfung ist kaum möglich. Es wird vorgeschlagen, im Studium ein Angebot von Beratungen, die auf unterschiedlichen Leistungsdaten aufbauen, einzuräumen, um damit zu einer Herausbildung realistischer Selbstkonzepte als Grundlage aktiver Passung von Fähig-

keiten und Anforderungen beizutragen. Das könnte auch die Prüflinge gegen nachteilige Wirkungen der Prüfungen immunisieren.

Exkurs I: Planung, Durchführung und Auswertung der empirischen Untersuchung

Die empirische Untersuchung, deren Ergebnisse hier berichtet werden, wurde 1970/71 im Rahmen eines größeren Projektes (s. Kapitel I) als Befragung von Prüflingen vor und nach Prüfungen am Psychologischen Institut der Universität Münster durchgeführt.

Übersicht: Im Teil 1 des Exkurses werden die grundlegenden Eigenschaften der Untersuchung im Hinblick auf die Absicht der empirischen Erkenntnisgewinnung begründet. Dazu gehören die Wahl der schriftlichen Befragung, die besondere Form der gestellten Fragen und die Art der Antwortvorgabe.

Die Planung und Durchführung der Untersuchung ist Gegenstand von Teil 2. Es wird über die Entwicklung der Fragebögen und die Durchführung der Untersuchungen berichtet.

Der abschließende dritte Teil hat die Probleme zum Inhalt, die mit der Verarbeitung der erhobenen Daten zusammenhängen. Über einen Bericht der statistischen Verarbeitung hinausgehend soll die Qualität der Daten durch Kontrolle der Auswirkungen der Selbstselektion und formaler Eigenschaften der Daten erkundet werden.

E I,1.: Methodische Anlage der Untersuchung

Die empirische Analyse der Rückmeldefunktion als Einfluß des Vollzugs der Prüfungen auf Selbstbewertung und Motivation ist Teil einer umfassenderen Untersuchung des Verhaltens und Erlebens von Prüflingen. An dieser Stelle soll eine Gesamtcharakteristik der Untersuchung mitgeteilt werden, die die thematischen Fragestellungen mit umfaßt. (Als thematische Fragestellungen werden die Operationalisierungen von Selbst-

bewertung und Motivation bezeichnet.)

Grundlegend gekennzeichnet ist das methodische Vorgehen durch die drei folgenden Merkmale: Die Untersuchung hat Erkundungscharakter und versteht sich nicht als hypothesentestende Empirie; die Daten werden sowohl vor als auch nach der Prüfung erhoben (prae-post-design); die Untersuchung bedient sich zur Datensammlung der schriftlichen Befragung.

Erkundungscharakter der Untersuchung

Die diesem Charakter entsprechende Absicht der Untersuchung ist die notwendige Folgerung aus dem in Kapitel III begründeten Verhältnis von Theorie und Empirie. Die Untersuchung geht davon aus, daß der Zustand theoretischer und empirischer Erkenntnis über Prüfungen für eine auf bestimmte Variablen und deren Zueinander eingegrenzte hypothesenorientierte Untersuchung nicht hinreicht. Aus Kapitel II dürfte deutlich geworden sein, daß ein solches "strenges" Vorgehen gar nicht das Ziel empirischer Betätigung im Rahmen hochschuldidaktischer Forschung sein sollte.

Der Erkundungscharakter der Untersuchung bringt es mit sich, daß durch geringe methodische Stringenz, fehlende Variablenkontrolle und Aspektvielfalt kaum die empirische Grundlage für die Entwicklung einer Theorie der Prüfung, die im strengen sozialwissenschaftlichen Sinn auf die Ergebnisse aufbaut, möglich sein wird. Er bietet aber den Vorteil, daß die verschiedenen Aspekte vielfältig die Empirie durchscheinen lassen, ihr in gewisser Weise eine freiere Ausdrucksmöglichkeit einräumen. Das ist zugleich die Voraussetzung dafür, daß die Prüflinge in der Untersuchung von ihnen als relevant erlebte Probleme wiedererkennen können, was nicht zuletzt eine effektive wie sinnvolle Grundlage der problemorientierten Kommunikation von Forschungsergebnissen als Mittel praktischer Problemlösung

ist. Für die thematischen Fragestellungen bedeutet es zudem die Chance, in vielen anderen Fragestellungen Hintergrundmaterial für die gegenstandsorientierte Deutung der Ergebnisse finden zu können.

Natürlich ist auch die Planung von Untersuchungen mit Erkundungscharakter nicht völlig ohne Hypothesen denkbar. Sie stellen in der vorliegenden Untersuchung Bereichsfestlegungen dar, die dem Rahmen der in Kapitel II angedeuteten Funktionentheorie der Prüfung entstammen. Grob angedeutet werden die folgenden Bereiche angesprochen: Vorbereitungsverhalten der Prüflinge; Erleben der erwarteten und der absolvierten Prüfungen; Einstellungen zu den Prüfungen. Es ist offensichtlich, daß Selbstbewertung und Motivation zu diesen Bereichen vielfältige Beziehungen aufweisen.

Verwendung der Methode schriftlicher Befragung

Die Darstellung der Methode rückt bestimmte Aspekte in den Vordergrund und verzichtet auf die aus ähnlichen Untersuchungen bekannte immer gleiche Wiederholung der Vor- bzw. Nachteile einzelner Fragetypen. Der Erkundungsuntersuchung, die einen durch Theorie und Empirie nur unzureichend vorstrukturierten Gegenstand unter natürlichen Feldbedingungen möglichst aspektreich erfassen soll, stehen aus der empirischen Sozialforschung die Methoden der mündlichen oder schriftlichen Befragung zur Auswahl (ERBSLÖH 1972; MANGOLD 1967; SCHEUCH 1973). Die empirisch sozialwissenschaftlichen Auswertungsmethoden setzen anders als erkundende Fallstudien umfangreiche Stichproben aus Datenkollektiven voraus. Der Umfang der Datenmengen soll Aussagen gestatten, die über individuelle Besonderheiten hinausgehen. Solche Datenmengen sind unter den gegebenen Bedingungen nur über Befragungen zu gewinnen.

Die Wahl der schriftlichen Befragung (Fragebogen)

gegenüber der mündlichen ist aus mehreren Gründen angezeigt. Die Situation vor der Prüfung und insbesondere die nach der Prüfung ist so beschaffen, daß eine mündliche Befragung in einem bestimmten zeitlichen Abstand von der Prüfung nur mit einer geringen Beteiligung von Prüflingen rechnen könnte. Es ist zu erwarten, daß die Beteiligung von Personenbedingungen beeinflußt wird, die gleichzeitig wesentliche Konstituentien des Verhaltens und Erlebens bezogen auf Prüfungen sind. Demgegenüber ist eine Teilnahme an der schriftlichen Befragung wahrscheinlicher und sogar dann noch möglich, wenn die Prüflinge Münster nach der Prüfung verlassen. Der geringere Aufwand läßt es außerdem zu, die Untersuchung auf verschiedene Fächer auszuweiten.

Die Begründungen sind nur dann hinreichend, wenn bei der gewählten Methode sichergestellt werden kann, daß die qualitative Struktur der gewonnenen Daten der Untersuchungsabsicht entspricht. Für die thematischen wie für die übrigen Fragestellungen hätte die mündliche Befragung kaum Vorteile gegenüber der schriftlichen erbracht. Die beabsichtigte Anwendung komplexerer statistischer Analysen fordert Daten in Form von geschlossenen Fragestellungen. Die mündliche Befragung würde deshalb für die Konstitution der Daten selbst nicht bedeutsam sein, sondern nur über einen Einfluß auf die Beantwortermotivation durch die Situation direkter Kommunikation wirken. Solche zusätzliche Motivation scheint überflüssig, weil durch die Aspektvielfalt in Verbindung mit der Berücksichtigung von Fragestellungen, die die Prüflinge selbst in den Voruntersuchungen (s. unten) vorschlagen konnten, die Motivation der Prüflinge zur "wahrheitsgemäßen" Beantwortung als gesichert gelten kann.

Anlage der Untersuchung als prae-post-design

Das Design der Untersuchung kann nach CAMPBELL (SCHWARZ 1970) als vorexperimentelle Anordnung der Struktur O_1 - X - O_2 bezeichnet werden. O_1 und O_2 stellen Meßwiederholungen dar, zwischen denen ein Prozeß liegt, der als zwischenzeitliches Geschehen charakterisiert wird. Die gleichzeitige Anwendung einer Vorher- und einer Nachhermessung ist notwendig, um die Wirkung auf die Prüfung und dort insbesondere auf die Noten einzugrenzen. Eine positive Beziehung zwischen Noten und Selbstbewertung nach der Prüfung zeigt erst dann einen Einfluß der Noten an, wenn diese Beziehung nicht schon vor der Prüfung und Notenvergabe bestand.

Die Anordnung wird bei SCHWARZ (1970) jedoch eher als ein Negativbeispiel angeführt, weil zu viele Störvariablen wirksam werden können. Als Störvariablen eines solchen Ansatzes kommen in Frage: 1. Die Ursachen für die Änderungen von O_1 nach O_2 können auf andere Prozesse als die in Augenschein genommenen zurückgehen; der zwischenzeitliche Prozeß ist in bezug auf die "wahre" Ursache nicht kontrollierbar. Kapitel III.I. geht, im Kontext einer weitergeführten theoretischen Gegenstandsbestimmung, auf dieses Kausalitätsproblem noch genauer ein.

2. Die übrigen aufgeführten Störvariablen, als die Reifung, Reaktivität der Messung in der Voruntersuchung, statistische Regression und Wechselwirkung von Voruntersuchung mit der Beeinflussung genannt werden, treten gegenüber der ersten so weit an Bedeutung zurück, daß eine Erörterung sich erübrigt.

Alternativen des Vorgehens, die mehr Kontrolle des zwischenzeitlichen Prozesses ermöglichen, sind nur schwer zu verwirklichen, weil sie mehr Bereitschaft von den Befragten fordern, als in dieser Situation zu erwarten ist. Gezielte Eingriffe experimenteller Art sind von vornherein ausgeschlossen, wenn man nicht den

Ernstcharakter der Situation aufgeben will.

Form der Fragen

Die Begründung der Form der Fragen betrifft neben vielen Fragestellungen, die nur am Rande in die Untersuchung einbezogen werden, auch die thematischen Fragen nach der Motivation und einige andere, die in Kapitel III.II. und Kapitel III.III. bearbeitet werden. Sie erfolgt deshalb ausführlich. Die thematische Frage nach der Selbstbewertung hat eine besondere Form und wird in Kapitel III.I. abgeleitet. Dort wird auf theoretischem Hintergrund auch der Frageninhalt der thematischen Fragen näher erörtert, der in diesem Exkurs nur angedeutet wird.

In den üblichen Methodenanweisungen wird die "Technik" der Fragengestaltung durch eine Reihe von beschreibenden Begriffen erläutert (ANGER 1969; SCHEUCH 1973). Die methodischen Anweisungen stützen sich dabei nur auf wenige und die immer gleichen, zumeist älteren empirischen Untersuchungen, die die einzelnen methodischen Probleme kaum systematisch in Angriff genommen haben. Die zur Beschreibung der Eigenschaften von Fragen oder Antworten oder beiden zugleich vorgeschlagenen Dimensionen - noch beliebter sind einfache Gegensatzpaare - verdecken teilweise die eigentlichen Probleme eher als sie zu lösen. Viele der getroffenen Entscheidungen müssen deshalb unbegründet bleiben oder gehen auf vage Spekulationen zurück. Das macht die ganze Misere der sozialwissenschaftlichen Methodenlehren in diesem Bereich deutlich, die mehr nur eine Folklore der Forschung darstellen, wie einer der Methodologen selbstkritisch feststellen muß (SCHEUCH 1973). Mangels Alternativen ist die Darstellung trotzdem auf die angebotenen Begriffe angewiesen. Es wird aber versucht, stellenweise

mehr, als die Begriffe es tun, zu differenzieren.

Die Fragen (der Endform) der Fragebögen besitzen nach ANGER (1969) die folgenden Merkmale: Sie sind durchweg standardisiert. Die Standardisierung geschieht in vielen Fragen durch das Antwortangebot in Form von Antwortskalen. (Bei ANGER werden sie Skalenfragen genannt.) Die Verwendung von Antwortskalen geht darauf zurück, daß das Spektrum der Auswertungsmöglichkeiten dann erweitert wird, wenn die Zusatzannahme der Intervallskaliertheit der erhaltenen Antworten gemacht wird. Bei vielen Fragen sind zudem Alternativantworten den Befragten wegen fehlender Differenzierungsmöglichkeit nicht zuzumuten.

Für die Konstruktion der Skalen sind mehrere Entscheidungen zu fällen, die die Stufenzahl und die Verankerung der Stufen betreffen. Dabei spielt mit herein, daß viele Fragen das Selbsterleben der Befragten ansprechen; sie verlangen Selbstbeurteilungen (self-ratings).

Die Befragung verwendet vorwiegend vierstufige Antwortskalen, die zumindest polar, weitgehend aber in allen Antwortstufen mit verbalen Ankern versehen sind. In der Frage der Skalengestaltung liegen wenige systematische Erfahrungen vor. Es fehlen gezielte Untersuchungen wie die von BENDIG (1953), der bei Selbstbeurteilungen gleichbleibende Zuverlässigkeiten bei ansteigender Zahl der Skalenstufen bis hin zu 9 Stufen nachweisen konnte. Er stellte zugleich fest, daß die verbale Verankerung der Skalenstufen die Zuverlässigkeit der Antworten erhöht.

In vielen Untersuchungen werden ungerade Stufenzahlen der Skalen bevorzugt. Es ist zu vermuten, daß dadurch eine bessere Annäherung der Antworten an die Normalverteilung und damit eine unbekümmertere Annahme der Intervallskalierung möglich werden soll. Trotz

dieser Überlegung schien eine gerade Stufenzahl angebracht. Die Zahl der Antwortstufen muß bei der großen Zahl beabsichtigter Fragen (Erkundungsuntersuchung) eher gering gehalten werden, um eine Antwortabgabe zu erleichtern. Damit die Gefahr einer Massierung der Antworten auf der mittleren Stufe umgangen wird, ist es sinnvoll, durch die gerade Stufenzahl einen gewissen Zwang auf die Verteilung der Urteile auszuüben.

Wenn möglich, sind alle Skalenstufen verbal verankert; in jedem Fall sind die beiden Randstufen verbal verankert. Die dominante Verwendung vierstufiger Antwortskalen bringt die Gefahr mit sich, daß die vier Antwortstufen sich im Laufe der Beantwortung von den konkreten verbalen Verankerungen ablösen und einen eher absoluten Sinn erhalten, der auf Konstanten des individuellen Urteilsprozesses zurückgeht.

Eine weitere Kennzeichnung von Fragen in der Alternative direkt/indirekt (ANGER 1969), die eher eine Dimension darstellt, gewinnt nur durch die Andeutung des Fragengegenstands und Frageninhalts einen Sinn. Gegenstand der Fragen mit Antwortskalen sind die einzelnen Fächer der mündlichen Prüfungen und die Prüfung insgesamt. Inhalte der Fragen sind Einschätzungen der einzelnen Prüfer oder Prüfungsfächer, Beschreibungen des eigenen Zustands, des Verhaltens und Erlebens während der Prüfungsvorbereitung und in der Prüfung. In den Fragen werden die Inhalte direkt angesprochen. Beispielsweise ist die Frage nach der Prüfungsangst formuliert: Beurteilen Sie bitte die Fächer danach, wieviel Prüfungsangst sie jetzt bei Ihnen erzeugen. Nach ANGER (1969) zeichnet eine direkte Frage aus, daß die Indexbildung bei Frager und Befragtem gleich ist. Hier wird deutlich, daß die Einteilung einer Frage nach ihrer Direktheit unterschiedliche Aspekte beinhaltet. Eine Frage ist zum einen dann direkt, wenn dem Antworter die Problemstellung der Frage klar ist,

d.h. wenn seine Vermutungen über den Inhalt in die Richtung weisen, die der Absicht des Fragestellers entspricht. Dabei kann der Untersucher den Inhalt von Frage (und Antwort) mit mehr theoretischem Hintergrund sehen als der Befragte. Ein zweiter Aspekt, von dem ersten nicht unabhängig, aber auch nicht mit ihm übereinstimmend, zielt die Möglichkeit an, ein Konstrukt wie etwa "Prüfungsangst" direkt, u.U. unter Aufnahme dieses Begriffs, zu erfragen, indem das Konstrukt selbst zum Inhalt der Frage wird. Zur Verdeutlichung soll ein weiteres Beispiel dienen. Direkt im ersten Sinn sind die Fragen in Intelligenztests, bei denen die Untersuchungsabsicht sowohl dem Untersucher (Frager) als auch dem Befragten klar ist, bei denen aber nur über die Anhäufung von vielen unterschiedlichen einzelnen Items eine verläßliche Gesamteinschätzung des Konstruktes Intelligenz erreicht wird. Unter dem zweiten Aspekt könnte man die Frage nach der Intelligenz unmittelbar an den Befragten weitergeben, indem man ihn seine eigene Intelligenz auf einer Selbstbeurteilungsskala einschätzen läßt. Die Fragen dieser Untersuchung sind als direkt im zweiten Sinn zu bezeichnen, was zugleich beinhaltet, daß auch eine Direktheit im ersten Sinn gegeben sein muß.

Dieses Vorgehen wird gestützt durch die kognitive Wende der Psychologie, die den Vorgängen der selbstbewußten Handlungssteuerung mehr Aufmerksamkeit widmet (MILLER et al. 1973) und auch die Motivationspsychologie erobert hat. Diese Entwicklung setzt voraus, daß im Individuum Erlebniskonzepte vorhanden sind, die prinzipiell eine Abfrage in solcher Direktheit ermöglichen. Das war auch bisher schon ohne expliziten Bezug auf die kognitive Wende in bestimmten Experimenten selbstverständlich. Beispielsweise wird in vielen Experimenten der Dissonanztheorie die abhängige Variable als Verhaltensbereitschaft durch eine einfache mehrstufige Skala operationalisiert. Ähnlich verfährt

BERLYNE (1974) zur empirischen Stützung seiner auf den Bereich des Wissenserwerbs spezifizierten Theorie der intrinsischen Motivation, indem er als Operationalisierung das auf einer Antwortskala durch Ankreuzen bekundete Ausmaß an Interesse verwendet. Auch im Bereich der Persönlichkeitsmessung wird die Anwendung direkter Befragung versucht (TAYLOR et al. 1972).

Auf den an den Indexbildungen der herkömmlichen Persönlichkeitspsychologie geschulten Beobachter wirken derartige Operationalisierungen unbedacht. Der Vorwurf kann jedoch der Persönlichkeitspsychologie zurückgegeben werden. Die Operationalisierung von Persönlichkeitseigenschaften durch dem Befragten uneinsichtige Persönlichkeitsskalen geht auf eine "Meß"theorie zurück, die annimmt, daß der Befragte mit seinen Antworten die Maxime verfolgt, sich nicht wahrheitsgemäß, sondern verfälschend gut darzustellen. Das mag zweifellos der Fall sein, wenn von den Ergebnissen entsprechender Persönlichkeitsmessungen Berechtigungen abhängig gemacht werden. In einer Untersuchung wie dieser jedoch ist das Problem möglicher Verfälschungen bei direkten Fragen differenzierter zu sehen. Die Verfälschung kann die folgenden Formen annehmen (die aber auch bei indirekten Fragen nicht immer auszuschließen sind):

1. Es besteht die Gefahr der Konfundierung der Antworten zu inhaltlich unterschiedlichen Fragen nicht nur durch die Verwendung von formal gleichen Skalen (s. oben), sondern auch dadurch, daß in jede Antwort auf eine direkte Frage eine allgemein wertende Stellungnahme zu Prüfungen einfließen kann. Neben diesen Beeinflussungen durch Einstellungsstereotypen können interpersonal unterschiedlich ausgeprägte Urteilstendenzen (BUNGARD & LÜCK 1974; INSKO 1967) die einzelnen Urteile intrapersonal einander angleichen.

2. Die direkte Art der Fragen ermöglicht den Befragten darüber hinaus eine Beantwortung, die, ge-

bunden an Vermutungen über den Zweck der Untersuchungen oder über die Art der Verwendung der Ergebnisse, nicht mehr eine einschätzende Bewertung des eigenen Verhaltens, Erlebens usw. darstellt, sondern als Mittel gezielter Mitteilungen an den Untersucher verwendet wird. Beispielsweise kann derjenige, der eine Veränderung von Prüfungen erreichen will, auch dann, wenn er selbst kaum Angst empfindet, von sich ein hohes Ausmaß an Prüfungsangst berichten, um damit ein Ergebnis der Untersuchung zu bewirken, das gegen Prüfungen spricht. Eine Kontrolle dieser Art der Verfälschung ist kaum möglich.

3. Die Antwortskalen können, auch wenn sie verbal verankert sind, was durch global wertende Begriffe wie "viel", "wenig" usw. geschieht, von verschiedenen Personen sehr unterschiedlich aufgefaßt werden. Das erhöht interindividuell die Unzuverlässigkeit der Antworten. Gerade dieser Einwand gilt aber ebenso für die Verwendung von Skalen bei indirekten Fragen.

Die gewählte Form der direkten Fragen bringt einen zusätzlichen Vorteil in praktischer Absicht ein. Die damit gewonnenen Ergebnisse sind unmittelbar in Beziehung zu den Fragestellungen darstellbar. Besonders mehrstufige Untersuchungen im Sinne der Aktionsforschung oder Untersuchungen, die eine zur Diskussion auffordernde unmittelbare Rückmeldung ihrer Ergebnisse verlangen, wie das in Kapitel II für die Hochschulforschung gefordert wurde, sind auf solche Verfahren angewiesen.

E I,2,: Planung und Durchführung der Untersuchung

Auch in diesem Teil werden nur solche Aspekte berichtet, die die Fragebögen in ihrer Gesamtheit betreffen; besondere methodische und auswertungstechnische Probleme einzelner Fragen und Auswertungsschritte sind in die Ergebnisteile integriert.

Entwicklung der Fragebögen

Dem beabsichtigten Design der Untersuchung entsprechend war die Entwicklung von zwei Fragebögen vorzubereiten. Einer davon, der Prüfungsvorbereitungsfragebogen (im folgenden: PVF) sollte vor der Prüfung, der andere, der Prüfungsfolgenfragebogen (im folgenden: PFF) sollte nach der Prüfung verteilt werden.

Die methodischen Anleitungen zur Entwicklung von Fragebögen (ANGER 1969; MANGOLD 1967) fordern eine sorgfältige mehrstufige Entwicklung. Dem wurde durch zwei Voruntersuchungen mit Vorformen der Fragebögen Rechnung getragen. Die Voruntersuchungen wurden mit Prüflingen im Fach Psychologie am Psychologischen Institut der Universität Münster durchgeführt.

In der ersten Voruntersuchung wurden die Vorform I des PVF und die Vorform I des PFF auf der Grundlage von mündlichen Befragungen von Prüflingen entwickelt.

Tabelle: E I. - 1

Versendung und Rücksendung der Vorform I des Prüfungsvorbereitungsfragebogens (PVF) und des Prüfungsfolgenfragebogens (PFF)

Prüfungstermine	PVF versendet	PVF zurück	PFF versendet	PFF zurück
Vordiplom				
Ende Wintersemester 69/70	10	8	-	-
Anfang Sommersemester 70	18	15	15	11
Gesamt	28	23	15	11
Hauptdiplom				
Ende Wintersemester 69/70	14	12	-	-
Anfang Sommersemester 70	19	14	18	12
Gesamt	33	26	18	12

Methodisches Vorgehen: Die Zahlen geben jeweils die Zahl der Exemplare an. Die Vorform des PFF wurde erst nach den ersten Auswertungen des PVF erstellt.

Die Vorform I des PVF bestand zum überwiegenden Teil aus offenen Fragen. Die Fragen selbst stimmten schon weitgehend mit denen der späteren Formen überein. Sie waren allerdings wesentlich breiter gestreut. In dem Fragebogen wurde an drei verschiedenen Stellen eine Kritik der Fragestellungen verlangt und um Alternativen gebeten. Antwortskalen waren nur in unbedeutender Anzahl eingearbeitet, die vorhandenen Skalen sahen viele (meist 7) Antwortstufen vor. Die thematischen Fragen waren bereits in geschlossener Form enthalten.

Die Versendung im Wintersemester 1969/70 erfolgte verspätet, so daß der Fragebogen teilweise erst im Verlauf der Prüfungen ausgefüllt wurde. Trotzdem wurde für die erste Phase der Durchführung auch eine inhaltliche Auswertung versucht. Erste Ergebnisse sind bei GERTMANN et al. (1971) mitgeteilt. Die Voruntersuchung bildete aber in erster Linie eine Entscheidungsgrundlage für die Fragenauswahl und Konstruktion der Antwortskalen. Die freien Antworten und die erwünschte Kritik gaben zusätzlich wertvolle Hinweise für Abänderungen und für Neuaufnahmen von Fragen.

Auf der Grundlage dieser ersten Voruntersuchung wurden die Formen II des PVF und PFF erstellt. Sie glichen schon weitgehend der Endform. Auch die Anordnung der Fragestellungen stimmte im wesentlichen mit der Endform überein. Gegenüber der Form I wurden fast ausschließlich geschlossene Fragen verwendet. Viele Fragen wurden mit Antwortskalen mit meist vier polar oder vollständig verbal verankerten Skalenstufen versehen. Bei den thematischen Fragen waren nur geringfügige Änderungen notwendig.

Da die zweite Voruntersuchung nur noch der Überprüfung der Verständlichkeit und Übersichtlichkeit der Fragenanordnung dienen sollte, konnte sie sich auf einen einzigen Prüfungstermin beschränken.

Tabelle: E I. - 2

Versendung und Rücksendung der Vorform II des PVF und des PFF

Prüfungstermin	PVF versendet	PVF zurück	PFF versendet	PFF zurück
Vordiplom				
Anfang Wintersemester 70/71	10	8	10	7
Hauptdiplom				
Anfang Wintersemester 70/71	6	6	6	6

Eine inhaltliche Auswertung der Ergebnisse war auf Grund der geringen Zahl der verteilten Fragebögen weder möglich noch Absicht. Die Fragebögen wurden allein auf Anmerkungen hin durchgesehen, und es wurde die Vollständigkeit der Angaben überprüft. Dabei stellte sich heraus, daß bei einzelnen Fragen etwas geänderte Anordnungen notwendig waren. So mußte bei den thematischen Fragen der Selbstbewertung und einer weiteren Frage der Fragengegenstand Gesamtprüfung optisch deutlicher herausgestellt werden, weil einige Prüflinge ihn übersahen.

Endform der Fragebögen

Im Anhang I sind die thematischen Fragestellungen und die übrigen, die im Rahmen dieser Arbeit ausgewertet werden, dargestellt. Es wird unterschieden zwischen thematischen Fragen (den abhängigen Variablen Selbstbewertung und Motivation) und den übrigen Fragen.

Die letzteren werden noch einmal unterteilt in solche, die sich auf die einzelnen Prüfungsfächer richten, und in solche, die das nicht tun. Die folgende Struktur der Kurzbezeichnungen wird in der Arbeit beibehalten; am Anfang der Kurzbezeichnung ist der Fragebogen, in dem die Frage enthalten ist, aufgeführt.

Beispiele:

Thematische Fragen:	PVF-SK.FÄCHER, PFF-MOT.BEL. usw.
Übrige Fächerfragen:	PVF - F 1. ATMOSPHÄRE, PVF - F 2. ANFORDERUNGEN usw.
Übrige Fragen:	PVF - I. NOTENSKALA usw.

Grob gesehen haben die Fragebögen die folgende Form: Der PVF beginnt mit der Befragung von Daten zum Studium, es folgen gemischt Fragen nach Verhalten, Erleben und Einschätzung der Prüfungsvorbereitung und Beurteilungen der einzelnen Prüfungsfächer in Form von Antwortskalen. Die thematischen Fragen sind, jeweils noch mit einer motivierenden Einleitung der Frage versehen (s. Anhang I), günstig in der Mitte des Fragebogens verstreut plaziert. Den Schluß bilden Fragen nach der Gruppenarbeit, einige Einstellungsskalen sowie abschließende Fragen zur Person des Befragten.

In einem zusätzlichen Anschreiben wird versucht, durch Hinweise auf die Untersuchungsabsicht (Notwendigkeit einer Reform der Prüfung, zu der die Daten der Fragebogenuntersuchungen beitragen sollen) die Motivation der befragten Studenten zur Mitarbeit an der Untersuchung wie zur wahrheitsgemäßen Mitteilung von Einschätzungen und Daten zu unterstützen. Die Reform gegenwärtiger Prüfungen ist nicht nur Grundlage der Untersuchung (s. Kapitel II), sondern auch übereinstim-

mendes Ziel der Studenten (TIMMERBRINK 1971).

Der PFF beginnt mit Fragen nach der Einstellung zur Prüfung, nach der Prüfung die beste Motivation zum Ausfüllen eines solchen Fragebogens. Die thematischen Fragen sind wiederum in der Mitte des Fragebogens verstreut eingebettet. Danach schließt sich ein weiterer Teil mit Einstellungsfragen an. Das Ende des Fragebogens bilden Fragen zu Verlauf und Erleben der einzelnen mündlichen Prüfungen. Diese Fragen, die neben Fragen mit Antwortskalen auch aus mehreren offenen Fragen sich zusammensetzen, werden auf die einzelnen Prüfungsfächer bezogen angeordnet, um die Rückerinnerung an die konkrete Prüfung zu intensivieren.

Gewinnung der Daten

Der PVF wurde zusammen mit der Einstellungsskala von TIMMERBRINK (1971) postalisch ungefähr zwei Wochen vor Beginn der mündlichen Prüfungen den Prüflingen zugestellt. Es wurden Freiumschläge zur Rücksendung der ausgefüllten Fragebögen beigefügt. Ungefähr eine Woche nach der Versendung wurde in den Fällen, in denen die Fragebögen ausgefüllt noch nicht zurückgesendet worden waren, schriftlich gemahnt. Teilweise waren auch persönliche Ansprachen der säumigen Ausfüller möglich.

Der PFF wurde entweder im Rahmen einer gleichzeitig durchgeführten Beobachtung mündlicher Prüfungen (SCHLIEHE & SEEGER 1974) nach der letzten mündlichen Prüfung verteilt, sonst (gleichfalls mit Rückantwortumschlägen) postalisch zugestellt. Nach einer Woche wurde gemahnt, teilweise wurde diese Mahnung dann noch einmal wiederholt.

Die Untersuchungsstichprobe setzt sich aus den Befragungen zu drei Prüfungsterminen zusammen. In die Stichprobe wurden Fragebögen des PVF dann nicht aufgenommen, wenn erkennbar war, daß die Ausfüllung erst

nach dem Absolvieren einiger oder aller mündlichen Prüfungen stattfand. Fragebögen des PFF wurden dann berücksichtigt, wenn die Ausfüllung bis zu vier Wochen nach der letzten mündlichen Prüfung erfolgte. Da die meisten Befragten, wie verlangt, das Ausfülldatum angaben, läßt sich außerdem mit Sicherheit feststellen, daß der verlangte Abstand einer Woche von der letzten mündlichen Prüfung weitgehend eingehalten wurde. Abstände bis zu vier Wochen wurden u.a. deshalb toleriert, weil sich in vielen Fällen die einzelnen Prüfungen selbst über einen längeren Zeitraum verteilen.

Der Umfang der Stichprobe wird, wie bei schriftlichen Befragungen nicht zu umgehen (ANGER 1969), durch ein größeres Ausmaß an Verweigerung der Teilnahme verringert. Eine zusätzliche Reduktion ergibt sich daraus, daß zur Auswertung nur Fragebögen gelangen, die für beide Befragungen zuordbar sind. In der folgenden Tabelle sind auch schon die Ausschlüsse wegen Nichteinhaltung des Zeitpunktes der Ausfüllung enthalten. Zusätzlich sind in der Tabelle auch die Rücksendungen in den beiden Vorformen berücksichtigt. Da bestimmte thematische Fragen identisch oder fast identisch auch in diesen Vorformen enthalten sind, sollen diese zur Vergrößerung des Stichprobenumfangs bei einigen Auswertungsschritten mit den Ergebnissen der eigentlichen Untersuchungen zu einer "weiten Stichprobe" zusammengefaßt werden. Das Verfahren ist nicht unproblematisch, weil die Fragenformulierung teilweise von der der Endform etwas abweicht und darüber hinaus die Einbettung der Fragen in den Fragebogen nicht gleich ist. Die Einbeziehung der Ergebnisse der Vorformen wird deshalb jeweils mit dem Hinweis: "Weite Stichprobe" versehen.

Tabelle: E I. - 3

Versendung und Rücksendung der Endform des PVF und PFF

(zusätzlich: Zusammenfassung zur "Weiten Stichprobe")

Prüfungstermine	versendet	PVF zurück	PFF zurück	zuordbar
Vordiplom				
Ende Wintersemester 70/71	16	14	12	12
Anfang Sommersemester 71	29	21	16	15
Ende Sommersemester 71	27	23	18	16
Gesamt	72	58	46	43
In Prozent von 72:		80,6	63,9	59,7
Vordiplom ("Weite Stichprobe")				
Anfang Sommersemester 70	18	15	11	11
Anfang Wintersemester 70/71	10	8	7	7
Gesamt ("Weite Stichprobe")	100	81	64	61
In Prozent von 100:		81	64	61

Fortsetzung der Tabelle s.S. 19

Tabelle: E I. - 3 (Fortsetzung)

Prüfungstermine	versendet	PVF zurück	PFF zurück	zuordbar
Hauptdiplom				
Ende Wintersemester 70/71	7	7	7	7
Anfang Sommersemester 71	30	20	16	14
Ende Sommersemester 71	13	10	8	8
Gesamt	50	37	31	29
In Prozent von 50:		74	62	58
Hauptdiplom ("Weite Stichprobe")				
Anfang Sommersemester 70	19	14	12	11
Anfang Wintersemester 70/71	6	6	6	6
Gesamt ("Weite Stichprobe")	75	57	49	46
In Prozent von 75:		76	65,3	61,3

Die Rücksendequoten sind zwar, gemessen an Erfahrungen mit Wiederholungsuntersuchungen, hoch, doch ist der Ausfall von ungefähr 40 % der Befragten immer noch erheblich. Die Auswirkungen dieser Selbstselektion sollen deshalb, soweit möglich, im folgenden Teil kontrolliert werden.

E I.3.: Datenqualität und Datenauswertung

Dieser Teil versucht, auch schon die Bestimmung solcher formalen Qualitäten der Daten einzubeziehen, die selbst Voraussetzungen der Auswertung darstellen oder die Gewichtigkeit der Ergebnisse beeinflussen können.

Besonderheiten der Daten

Bei den Beantwortungen einzelner Fragen ergeben sich einige Probleme durch unvorschriftsmäßige oder fehlende Antworten.

1. Bei ungenauen Ankreuzungen von Antwortskalen werden die Skalenpositionen gewählt, denen das gesetzte Kreuz am nächsten steht. In einigen wenigen Fällen, die danach nicht entscheidbar sind, oder bei Doppelankreuzungen wird nach Zufall zugeordnet.

2. In angenähert 10 % aller Fragebögen fehlt eine Angabe für die Notenerwartung in der Gesamtprüfung (PVF - I. NOTENERW.GESAMT), in weniger als 5 % der Fälle fehlt eine entsprechende Angabe für den Gesamtwert des Selbstkonzepts (PVF/PFF-SK.GESAMT). Wenn aus den Notenerwartungen und Selbstkonzepten für die Einzelfächer deutlich eine Gesamttendenz herauszulesen ist, wird daraus der Gesamtwert geschätzt.

Mit Ausnahme der etwas häufigeren Änderungen unter Punkt 2 beschränken sich die übrigen Änderungen auf ein oder zwei Beantworter pro betreffende Frage.

Eine weitere Entscheidung wird notwendig, weil durch Krankheit eines Prüfers in einem Prüfungstermin die Prüflinge von einem anderen Prüfer als von dem, auf den sie sich vorbereitet hatten, geprüft wurden; die Ergebnisse werden in diesem Fall dem Ersatzprüfer zugerechnet, weil die Prüfung selbst als die wesentlichere Einflußquelle angesehen wird. Im Vordiplom betrifft das im Fach Persönlichkeitspsychologie 10 Prüfungen, die sämtlich von demselben anderen Prüfer übernommen wurden. Im Hauptdiplom wurden im Fach Sozialpsychologie 7 Prüfungen durch einen anderen Prüfer übernommen, und im Fach Pädagogische Psychologie war das bei 3 Prüfungen der Fall.

Dort, wo die Angabe eines Prüfers zwischen PVF und PFF nicht übereinstimmt, erfolgt keine Zurechnung

zu einem bestimmten Prüfer. Das kommt aber nur im Vordiplom in einigen wenigen Einzelprüfungen vor.

Eine Besonderheit bei den Nebenfächern im Vordiplom entsteht dadurch, daß bestimmten Prüflingen Nebenfachprüfungen aus anderen vorher schon absolvierten Prüfungen anerkannt werden. Antworten für diese Prüfungen werden nicht berücksichtigt.

Die Noten, die die Prüflinge in den Prüfungen erhalten haben, werden nicht erfragt, sondern direkt aus den Prüfungsakten übertragen.

Datenverarbeitung

Die Antworten wurden kodiert und auf Lochkarten übertragen. Die Rechenarbeiten wurden am Rechenzentrum der Universität Münster unter Verwendung von Programmen der Programmsammlungen SPSS (NIE et al. 1975), DRZ (GEBHARDT 1969), B. RAMING (unveröffentlicht) und von eigenen für spezielle Zwecke erstellten Programmen durchgeführt.

Anwendung statistischer Verfahren

Die Verwendung von statistischen Verfahren in deskriptiver, mehr noch die in inferenzstatistischer Absicht bringt in einer derartigen Untersuchung einige Probleme mit sich. Außer den für jedes Verfahren spezifischen Problemen sind drei allgemeine zu nennen: Skalenniveau und Verteilungseigenschaften der verwendeten Daten; Grundlagen der statistischen Verallgemeinerung; vielfache Anwendung von Signifikanztests in ein und derselben Stichprobe.

Die unterschiedlichen statistischen Verfahren setzen ein bestimmtes Skalenniveau der Daten voraus, das gewöhnlich als Skalentyp definiert wird. Entsprechend ist eine Entscheidung über das Skalenniveau durch

Zuordnung der Daten zu einem Skalentyp zu treffen. Während sie zwischen Nominalskala und Ordinalskala selten schwerfällt, ist sie zwischen Intervallskala und den ersteren problematischer. Sie ist zugleich bedeutsamer, weil die für Intervallskalen entwickelten parametrischen Verfahren erweiterte Verfahrensmöglichkeiten eröffnen. Die parametrischen Verfahren setzen in der Regel zusätzlich eine Normalverteilung der entsprechenden Meßwerte in der Population voraus. Die Feststellung, ob Daten intervallskaliert vorliegen, ist in direktem Zugriff nicht möglich. Deshalb wird üblicherweise mit einer Überprüfung der Normalverteilung vorliebgenommen, die von der Annahme begleitet wird, daß die entsprechenden Merkmale dann auch intervallskaliert sind, wenn in der Stichprobe ihre Normalverteilung nicht widerlegt werden kann.

Zur empirischen Prüfung der Normalverteiltheit werden die Antworten mit dem Kolmogoroff-Smirnoff-Test auf Güte der Anpassung (SIEGEL 1956) getestet. Das Ergebnis dieser Überprüfung, die für die thematischen Variablen teilweise im einzelnen in Kapitel III.I. dargestellt wird, lautet: Von über 200 überprüften Verteilungen der Antworten auf Skalen weichen nur einige wenige bedeutsam von der Normalverteilung ab. Zwar sind mehrere Verteilungen deutlich erkennbar links- oder rechtsschief, eine statistische Bedeutsamkeit aber erreichen die Abweichungen nicht. Die statistisch bedeutsamen Abweichungen liegen zudem bei Fragen vor, die im Rahmen dieser Arbeit kaum Beachtung finden.

So scheint, von den Verteilungsvoraussetzungen her gesehen, eine Anwendung parametrischer Verfahren gerechtfertigt. Wenn sie bevorzugt angewendet werden (auch dort, wo wie bei den Noten die Voraussetzungen fragwürdig sind), hat das aber noch einen anderen Grund. Die Fehler, die bei Anwendung parametrischer Verfahren wie t-Tests dadurch entstehen, daß die Daten

nicht normalverteilt sind, haben sich nicht als sehr gravierend herausgestellt. Vorwiegend bewirken sie eine konservativere Testung. Die ist aber auch mit vielen parameterfreien Verfahren verbunden, die die enthaltene Information nur zum Teil ausnutzen. Darüber hinaus fehlen gerade solche statistischen Verfahren, die die Information von Daten ausnutzen, die auf einer homomorphen Ordinalskala (ATHANASIADIS 1974) gemessen werden. Die Verwendung von statistischen Methoden, die bei Rangskalen angebracht sind, würde bei den vorliegenden Daten einen derart großen Anteil von Rangbindungen ergeben, daß die Bindungskorrekturen, die nur Annäherungsverfahren darstellen, das kaum ausgleichen dürften. (Allerdings lassen die statistischen Texte, auch neuerlich LIENERT (1973) es an genaueren Hinweisen fehlen.)

An einigen Stellen werden parametrische und parameterfreie Verfahren gleichzeitig angewendet, was allerdings zusätzlichen Interpretationsaufwand erfordert.

Die inferenzstatistische Bewertung der Ergebnisse erfolgt ausschließlich auf der Grundlage zweiseitiger Testung. Das entspricht dem Charakter der Erkundungsuntersuchung. Außerdem sind selten genügend theoretisch oder empirisch begründete Hinweise auf gerichtete Hypothesen möglich.

Wenn in einer Untersuchung inferenzstatistische Methoden angewendet werden, stellt das die Frage nach dem Verhältnis von Stichprobe und Population. Bei Felderhebungen wie diesen, die mit einer Gruppe von Personen unter zeitlich und örtlich besonderen Bedingungen arbeiten, gibt es im strengen Sinn keine Stichprobe, die als zufällige Auswahl aus der Population gelten kann. Insofern ist auch kein inferenzstatistisch abgesicherter Schluß von der Stichprobe auf die Population möglich.

Wenn trotzdem solche Verfahren zur Entscheidung über die Interpretierbarkeit eines Ergebnisses herangezogen werden, hat das weiterhin die Bedeutung, gegen zufällige Ergebnisse abzusichern. Nur ist diese Absicherung jetzt in Analogie zu den "Ceteris-paribus-Klauseln" der modellkonstruktiven Wissenschaften zu sehen (ALBERT 1965). Die Argumentationsweise wäre die folgende: Unter den gleichen Bedingungen wie denen der Untersuchung würde bei einem Austauschen der Ereignisse (Personen) unter Zufallsvariation mit einer der Wahrscheinlichkeitsangabe des statistischen Schlusses entsprechenden Wiederholbarkeit des Ergebnisses zu rechnen sein. Die Feststellung der Gleichheit der örtlichen und zeitlichen Bedingungen bleibt dann dem "Abnehmer" der Untersuchungsergebnisse überlassen. Zur Feststellung der Gleichheit der Bedingungen braucht der Abnehmer eine Fülle von Informationen über die Umstände der Untersuchung, die Besonderheiten der Personen usw. (s. auch HOLZKAMP (1966) für die entsprechende Problemstellung in der psychologischen Diagnostik). In Kapitel III.I. wird versucht, diese Informationen bereitzustellen.

Ein Problem, das mehr oder weniger alle erkundenden Fragebogenerhebungen dieser Art betrifft, ist die Datenfülle. Es gibt Versuche, diese Datenfülle durch dimensionsanalytische Verfahren (wie die Faktorenanalyse) zu reduzieren; sie sind aber aus zwei Gründen in der vorliegenden Untersuchung nur begrenzt anwendbar. 1. Einzelne Fragen stehen als abhängige Variablen im Mittelpunkt des theoretischen Interesses. 2. Der Fragebogen ist nicht besonders auf eine solche Analyse hin angelegt.

Viele unterschiedliche Fragen aber legen die, auch probierende, Auswertung über eine Vielzahl von statistischen Kennwerten nahe. Dadurch wird den Modellen, die einer Bestimmung statistischer Signifikanz

zugrunde liegen, nicht entsprochen. Die Modelle setzen die Unabhängigkeit der Kennwerte voraus. Die Kennwerte sind jedoch über die Eigenarten der Personen und Situationen miteinander in komplexer Art verbunden. Dieser Einwand wird häufig in dem Sinn mißverstanden, daß durch solche mehrfache Kennwertbildung nach den Gesetzen des Zufalls schon "signifikante" Ergebnisse sich zufällig ergeben müßten. Das gilt für Kennwerte, die jeweils an voneinander unabhängigen Zufallsstichproben gewonnen werden, nicht aber für mehrfache Kennwertbildung in derselben Stichprobe, wenngleich die Richtung möglicher Fehlurteile dieselbe ist: Die Signifikanz von Befunden wird eher überschätzt. Konkrete Korrekturen aber sind dafür nicht möglich.

Auswirkungen der Selbstselektion

Ein erster Schritt zur Absicherung der formalen Güte der Ergebnisse besteht in der Überprüfung der Auswirkungen der erheblichen Selbstselektion, die in Teil 2 festgestellt wurde. Als Grundlage des Vergleichs von Beantwortern und Nichtbeantwortern bieten sich die Daten an, die aus der Auswertung der Prüfungsakten, denen die Noten entnommen wurden, zur Verfügung stehen. Es handelt sich auf der einen Seite um die Personencharakteristika Alter und Geschlecht, auf der anderen Seite um die in den Prüfungen erteilten Noten sowie um die Wahl der Prüfungsordnung (s. Kapitel III.I.).

Ein direkter Schluß von dem Vergleich der Beantworter mit den Nichtbeantwortern auf die Wahrscheinlichkeit möglicher Verzerrungen der Ergebnisse ist noch am besten an Hand der Noten möglich. Aber auch dort gilt, daß die einzelnen Auswertungen der Arbeit von möglichen Unterschieden zwischen Beantwortern und Nichtbeantwortern unbeeinflußt bleiben können, während auch bei hier festgestellter Gleichheit der beiden Gruppen komplexe Einflüsse der Selbstselektion

auf funktionale Beziehungen zwischen bestimmten Variablen bestehen können.

Die Beantworter und Nichtbeantworter unterscheiden sich kaum auffällig in Alter und Geschlecht. Auch die gewählte Prüfungsordnung verteilt sich nicht statistisch bedeutsam unterschiedlich auf die beiden Gruppen.

Tabelle: E I.- 4

Vergleich von Beantwortern und Nichtbeantwortern in Alter, Geschlecht und Wahl der Prüfungsordnung

Variable	Beantworter	Nichtbeantw.	Wahrscheinlichkeit zuf. Abw.
Vordiplom			
Alter			
unter 25	25	15	p = .77
gleich/über 25	18	14	
Geschlecht			
männlich	26	16	p = .84
weiblich	17	13	
Prüfungsordnung			
APO	25	13	p = .38
NPO	18	16	
Hauptdiplom			
Alter			
unter 27	15	12	p = .93
gleich/über 27	14	9	
Geschlecht			
männlich	16	12	p = 1.00
weiblich	13	9	
Prüfungsordnung			
APO	7	4	p = .94
NPO	22	17	

Methodisches Vorgehen: Die Vergleiche beider Gruppen erfolgen mit dem Fisher-Yates-Test (SIEGEL 1956). Dieser Vergleich ist eine strenge Form der Prüfung.

Mit Ausnahme der Entscheidung für eine der beiden Formen der Prüfungsordnung im Vordiplom sind die beiden Gruppen einander weitgehend ähnlich. Aber auch der Unterschied der gewählten Prüfungsordnung ist nicht annähernd signifikant.

Zentraler ist, wie erwähnt, die Überprüfung der Gleichheit der Notenverteilung in den beiden Gruppen der Beantworter und Nichtbeantworter. Der zentrale Stellenwert ergibt sich unmittelbar daraus, daß die Noten die wesentliche unabhängige Variable der Untersuchung sind. Die Gleichheit beider Gruppen wird für die Gesamtnoten, für die Einzelnoten und über alle Noten der einzelnen mündlichen Prüfungen hinweg (s.Exkurs II) untersucht.

Tabelle: E I. - 5

Vergleich von Beantwortern und Nichtbeantwortern in den Noten der Prüfung

Note in:	Beantworter 1	Beantworter >1	Nichtbeantw. 1	Nichtbeantw. >1	Wahrscheinlichkeit zuf. Abweichung
Vordiplom					
Schriftl.Hausarb. (nur APO)	12	13	11	2	p = .06
Gesamtnote	25	18	18	11	p = .94
Ausdr./Method.	18	25	19	10	p = .08
Allg.Psychol.	33	10	23	6	p = 1.00
Entw.Psychol.	34	9	19	10	p = .31
Pers.Psychol.	28	15	13	16	p = .14
Physiologie	12	28	10	17	p = .73
Wahlfach I	16	21	15	9	p = .23
Wahlfach II	17	18	13	11	p = .88
Über alle Fächer	158	126	112	79	p = .58 (?)
Über alle psychol.Fächer	113	59	74	42	p = .83 (?)
Über alle nichtpsychol.Fächer	45	67	38	37	p = .21 (?)

Tabelle: E I. - 5 (Fortsetzung)

Vergleich von Beantwortern und Nichtbeantwortern in den Noten der Prüfung

Note in:	Beantworter 1	Beantworter >1	Nichtbeantw. 1	Nichtbeantw. >1	Wahrscheinlichkeit zuf. Abweichung
Hauptdiplom					
Gesamtnote	25	4	7	14	p = .00
Diagnostik	22	7	7	14	p = .01
Sozialpsychol.	27	2	13	8	p = .02
Päd.Psychol.	22	7	14	7	p = .69
Klin.Psychol.	26	3	13	8	p = .05
Angew.Psychol.	22	7	11	10	p = .15
Über-alle-Fächer	119	26	58	47	p = .00 (?)

Methodisches Vorgehen: Wie für Tabelle: E I. - 4 erfolgen die Vergleiche der beiden Gruppen mit dem Fisher-Yates-Test. Die Noten werden in zwei Klassen: "Note 1" und "Note schlechter als 1" aufgeteilt. Eine Erläuterung zu den einzelnen Fächern der Prüfung erfolgt in Kapitel III.I.

Im Vordiplom sind nur bei der schriftlichen Hausarbeit und im Fach Ausdr./Method. deutliche Abweichungen festzustellen, die aber nicht signifikant sind. In beiden Fällen haben die Beantworter des Fragebogens die schlechteren Noten. In den anderen Fächern weisen nicht einmal die Tendenzen in die gleiche Richtung, so daß sich alle Noten der mündlichen Prüfungen wie auch die Gesamtnoten in beiden Stichproben nicht mehr deutlich unterschiedlich verteilen.

Insgesamt kann deshalb im Vordiplom die durch Selbstselektion entstandene Gruppe der Beantworter als einigermaßen repräsentativ für die Gesamtgruppe angesehen werden.

Für die Prüflinge des HD gilt das nicht. Dort haben signifikant mehr Prüflinge mit guten Noten die Fragebögen beantwortet als nicht beantwortet. Die Unterschiede der beiden Gruppen sind nur in zwei Fächern der mündlichen Prüfungen, in Pädagogischer Psychologie und Angewandter Psychologie, nicht signifikant. Ergebnisse, die aus den Prüfungen des HD, besonders in Zusammenhang mit den Noten gewonnen werden, können deshalb nicht einmal Repräsentativität für die Gesamtgruppe der Prüflinge im HD in den drei Prüfungsterminen am Psychologischen Institut der Universität Münster beanspruchen.

Formale Qualität der Daten

Eine Feststellung der formalen Qualität der Daten erfolgt zugleich mit der Darstellung der Ergebnisse. Jedes theoretisch einordbare Ergebnis ist ein Beitrag zur Konstruktvalidität der Antworten. Die Validierung der Antworten in jener einfachen Form, die selbst die Testpsychologie abzulegen beginnt, ist nicht angemessen. Dafür erhält eine Kontrolle der Zuverlässigkeit eine aufgewertete Bedeutung, die auf ihre von der klassischen Testtheorie formulierte Beziehung zur Validität zurückgeht. Die Zuverlässigkeit ist im Hinblick darauf zu bestimmen, daß das Ausbleiben oder die Geringfügigkeit bestimmter empirisch feststellbarer Beziehungen zwischen den Fragen nicht schon durch deren Unzuverlässigkeit begründet werden kann. Die Zuverlässigkeit kann durch zwei Ansätze bestimmt werden, die nicht voneinander unabhängig sind. Sie ist sowohl feststellbar an der Übereinstimmung der Antworten zu gleichen Fragen im PVF und PFF als auch innerhalb eines Fragebogens durch die Übereinstimmung zwischen Gesamtwerten, die aus den Einzelfächerantworten berechenbar sind, mit solchen Gesamtwerten, die von den Antwortern direkt eingeschätzt

wurden.

Für das erste Verfahren (Übereinstimmung der Antworten von PVF und PFF) stehen nur insgesamt vier der Fragen zur Verfügung, davon sind drei thematische Fragen. Die erhaltenen Werte stellen eher Unterschätzungen der Zuverlässigkeiten der Antworten dar, da mit einiger Sicherheit systematische Verschiebungen in den Antworten durch den Einfluß der Prüfung und der dazwischenliegenden Zeitspanne anzunehmen sind.

Tabelle: E I. - 6

Korrelationen der Antworten im PVF und PFF
Einzelfächer und mittlere/Gesamtkorrelationen

Fragestellung	Ausdr./Method.	Allg.Psychol.	Entw.Psychol.	Pers.Psychol.	Physiologie	Wahlfach I	Wahlfach II	Mittl. Korr. Fächer	Mittl. Korr. Fach-Prüf.-Komb.	Korr. über alle Fächer
Vordiplom										
PVF/PFF-SK.FÄCHER	.64	.19	.52	.54	.43	.58	.61	.51	.48	.51
PVF/PFF-MOT.W.A.	.73	.59	.67	.72	.62	.70	.40	.64	.56	.72
PVF-F 6/PFF-F 9 Wichtigkeit	.67	.10	.47	.16	.54	.68	.54	.48	x	.60
PVF/PFF-MOT.BEL.	.33	.32	.36	.63	.46	.70	.31	.46	.46	.53

x = Wert fehlt, weil diese Berechnung nur für die thematischen Fragen vorgenommen wurde (s. auch Forts.d.Tab.)

Fortsetzung der Tab. s.S. 31

Tabelle: E I. - 6 (Fortsetzung)

Korrelationen der Antworten im PVF und PFF
Einzelfächer und mittlere/Gesamtkorrelationen

Fragestellung	Diagnostik	Sozialpsychol.	Päd.Psychol.	Klin.Psychol.	Angew.Psychol.	Mittl. Korr. Fächer	Mittl. Korr. Fach-Prüf.-Komb.	Korr. über alle Fächer
Hauptdiplom								
PVF/PFF-SK.FÄCHER	.49	.59	.07	.56	.73	.51	.45	.50
PVF/PFF-MOT.W.A.	.83	.72	.57	.69	.80	.73	.71	.75
PVF-F 6/PFF-F 9 WICHTIGKEIT	.66	.38	.30	.61	.58	.52	x	.57
PVF/PFF-MOT.BEL.	.54	.44	.52	.54	.59	.53	.51	.57

Methodisches Vorgehen: Die Korrelationen werden als Produkt-Moment-Korrelationen berechnet. In drei zusätzlichen Spalten sind mittlere Korrelationen dargestellt. In der ersten Spalte wird die mittlere Korrelation der fächerweise berechneten Korrelationen wiedergegeben. Sie wird durch arithmetische Mittelung der z-transformierten Werte gewonnen. In der zweiten Spalte findet sich der Mittelwert der Korrelationen, die in den einzelnen "Fach-Prüfer-Kombinationen" (s. Exkurs II) bestehen. Auch hier werden z-transformierte Werte gemittelt, in diesem Fall nach der Personenzahl gewichtet. In der dritten Spalte sind die Korrelationen über alle Antworten hinweg (s. Exkurs II) wiedergegeben.

Bei den Korrelationen der Einzelfächerantworten sind insbesondere die höchsten erreichten Werte interessant, weil angenommen werden kann, daß sie eher ein

Maß der Zuverlässigkeit darstellen, während die niedrigeren Korrelationen durch systematische Einflüsse verringert werden. Dies berücksichtigend kann auch mit Blick auf die mittleren Interkorrelationen davon ausgegangen werden, daß die Antworten durchaus zuverlässig sind und einen weiten Spielraum für inhaltlich bedingte Zusammenhänge bereitstellen.

Eine weitere Methode zur Bestimmung der Zuverlässigkeit der Antwortskalen, die innerhalb einer Befragung berechnet werden kann und dadurch den die Zuverlässigkeit senkenden Einfluß von zwischenzeitlichen Prozessen ausschaltet, besteht darin, die Übereinstimmung zwischen den aus den Antworten für Einzelfächer berechneten Gesamtwerten und den von den Prüflingen selbst eingeschätzten Gesamtwerten zu berechnen. Nur die Frage nach dem Selbstkonzept (PVF/PFF-SK.FÄCHER) ermöglicht ein solches Vorgehen. Dabei wird aus den Antworten zu den Einzelfächern unter Übergehen der methodischen Bedenken ein arithmetischer Mittelwert gebildet. Die Mittelwertbildung wird nur dann vorgenommen, wenn Einschätzungen für mehr als vier Einzelfächer vorliegen.

Tabelle: E I. - 7

Korrelationen zwischen errechneten und von den Prüflingen selbst eingeschätzten Gesamtwerten in PVF/PFF-SK.FÄCHER mit PVF/PFF-SK.GESAMT

	PVF		PFF	
Vordiplom	.72	N = 36	.91	N = 30
Hauptdiplom	.81	N = 27	.86	N = 28

Methodisches Vorgehen: Die Korrelationen werden als Produkt-Moment-Korrelationen berechnet. Das N gibt die Zahl der in die Auswertung einbezogenen Personen an, die besonders im VD durch die Anerkennung von

Nebenfachprüfungen verringert wird.

Die erhaltenen Werte sind sehr hoch. Sie sollten aber nicht überbewertet werden, weil die Einschätzung der Gesamtwerte unmittelbar im Anschluß an die Einschätzungen der Einzelfächer verlangt wird (s. Anhang I) und das rein optisch eine Mittelung nahelegt.

Im Teil 1 dieses Exkurses wurde erörtert, ob die Verwendung gleichartiger Antwortskalen das für den Frageninhalt spezifische Antwortverhalten einenge und sich im Laufe der Beantwortung von den konkreten Bedeutungen der Skalenstufen löse. Einer Überprüfung sollen die Daten der folgenden Tabelle dienen. In ihr sind alle mit vierstufigen Antwortskalen ausgestatteten Fragen, die sich auf die Einzelfächer der mündlichen Prüfungen beziehen, zusammengefaßt.

Tabelle: E I. - 8

Mittelwerte der pro Beantworter gemittelten Einzelfachantworten im PVF und PFF
(nur 4-stufige Antwortskalen)

Frage	Vordiplom arithm.Mittel	Hauptdiplom arithm.Mittel
PVF - F 1. ATMOSPHÄRE	2.78	2.77
PVF - F 2. ANFORDERUNGEN	2.18	2.14
PVF - MOT.BEL.	2.37	2.20
PVF - F 3. FESTLEGUNG	2.11	2.14
PVF - F 4. ANNEHMLICHKEIT	2.30	2.32
PVF - F 5. RELEVANZ	1.96	1.81
PVF - F 6. WICHTIGKEIT	2.41	2.13
PVF - F 7. PRÜFUNGSANGST	2.85	2.79
PVF - MOT.W.A.	2.88	2.37
PFF - MOT.W.A.	2.84	2.32
PFF - F 9. WICHTIGKEIT	2.45	2.17
PFF - MOT.BEL.	2.48	2.24
PFF - F 12. RÜCKM.NEG.	3.25	3.30
PFF - F 13. RÜCKM.POS.	3.29	3.30
PFF - F 14. KONGRUENZ.VORB.	1.96	1.89
PFF - F 15. KONGRUENZ.STUD.	2.54	2.54
PFF - F 16. ANPASSUNG	2.04	1.82
PFF - F 17. ATMOSPHÄRE	3.09	3.13
PFF - F 18. ANFORDERUNGEN	2.39	2.60
PFF - F 19. PRÜF.INSGESAMT	2.11	2.01

Methodisches Vorgehen: Die Mittelwerte werden über alle Antworten der Personen und Prüfungsfächer hinweg berechnet. Sie sind ein Maß des auf den Inhalt der Frage (und Antwortskala) zugeschnittenen Gebrauchs der Antwortskalen.

Die Auswertung ist rein beschreibend und beschränkt sich auf wenige Details der Tabelle. Sowohl im VD als auch im HD werden große Unterschiede zwischen den einzelnen Fragen deutlich, die für einen inhaltsspezifischen Gebrauch der Skalen sprechen. Diese Annahme wird bestätigt durch die Gleichheit der Mittelwerte gleichlautender Fragen (PVF/PFF-MOT.BEL.; PVF/PFF-MOT.W.A.; PVF-F 6./PFF-F 9.WICHTIGKEIT); beim VD und beim HD weichen die Mittelwerte nur geringfügig voneinander ab, obwohl die Fragen im PVF und PFF in einem ganz unterschiedlichen Kontext stehen. So wäre bei stark kontextabhängigem Skalengebrauch kaum zu erwarten, daß die Frage PFF-MOT.W.A., die dort die erste Frage mit Antwortskala darstellt, weitgehend identische Mittelwerte mit der inhaltsgleichen Frage PVF-MOT.W.A. aufweist, die dort als letzte an eine lange Reihe von Fragen mit formal gleichen Antwortskalen anschließt.

Erscheinen von daher die auf den Antwortskalen gegebenen Antworten als eher an einem inhaltsbedingten absoluten Bezugssystem orientiert, läßt eine andere Besonderheit der Ergebnisse wiederum Zweifel an der Qualität der Antworten zu. Der Vergleich der Mittelwerte der Antworten zwischen VD und HD ergibt eine auffallende Gleichheit dieser Mittelwerte bei gleichen Fragen, obwohl die einzelnen Fächer nichts miteinander gemein haben. Die Deutung dieses Ergebnisses kann unterschiedlich erfolgen. Nimmt man an, daß die Summe der Einzelfächer im HD wie im VD einen objektiv in etwa gleichen Gegenstand konstituiert, dann wäre die Gleichheit der Mittelwerte von daher verständlich.

Diese Annahme aber kann bezweifelt werden. Dann muß die Gleichheit der Mittelwerte auf einen personenspezifisch stabilisierten Gebrauch der Skala zurückgeführt werden, der zwar inhaltsspezifisch, nicht aber gegenstandsspezifisch ist. Für die Auswertung ist diese letzte Erörterung nicht von unmittelbarer Bedeutung, da sie im wesentlichen von der Varianz zwischen Personen und nicht von der Varianz zwischen Inhalten bzw. Gegenständen ausgeht.

Exkurs II: Auswertungsebenen der Antworten zu Einzelfächern

Jeder Prüfling hat im Fragebogen die Fächer der mündlichen Prüfungen nach einer Vielzahl von Aspekten zu beurteilen (s. Anhang I: Fächerfragen). In einem Prüfungsfach prüfen meist mehrere Prüfer. Beides zusammen eröffnet verschiedene Auswertungsmöglichkeiten, die im folgenden als Auswertungsebenen bezeichnet werden. Um die Übersichtlichkeit der Darstellung zu gewährleisten und die einzelnen Schritte der Auswertung besser vergleichbar zu machen, soll der Versuch unternommen werden, die Zahl der Ebenen generell zu reduzieren. Die in jedem empirischen Schritt unterschiedliche Bestimmung der Auswertungsebenen je nachdem, auf welcher Ebene die klarsten Ergebnisse vorliegen, ist statistisch problematisch, weil dadurch Zufallseinflüssen zu viel Raum gegeben würde. Die generelle Festlegung der Auswertungsebenen im voraus schließt aber nicht aus, daß bei einzelnen Schritten der Auswertung trotzdem andere als die einmal festgelegten verwendet werden.

Die Auswahl orientiert sich an den Bedingungen, die auf den verschiedenen Ebenen als Grundlagen der Beziehung zwischen den Daten erfaßt werden. So ist die Darstellung zugleich ein Beitrag zur Unterscheidung von möglichen Einflußquellen.

Kennwerte (wie z.B. Mittelwerte und Korrelationen) können auf den folgenden Ebenen berechnet werden (Spalte 1 der Tabelle: E II. - 1):

Tabelle: E II. - 1

Ebenen der Auswertung und mögliche Bedingungen von Ergebnissen

Ebene der Auswertung	I Auswertung innerhalb	II Auswertung zwischen Kennwerten
	Bedingungen 1. Art	Bedingungen 2. Art
Fach-Prüfer-Kombination	Individuum/Situation	Prüfer Fach Prüfungsart (VD/HD)
Einzelfach	Individuum/Situation Prüfer	Fach Prüfungsart
Über-alle-Fächer	Individuum/Situation Prüfer Fach	Prüfungsart
Mittelwerte der Fach-Prüfer-Komb.	Prüfer Fach	Prüfungsart
Mittelwerte der Einzelfächer	Fach	Prüfungsart
	Kontingenzen	Parameterunterschiede

Statistisches Paradigma

1. Kennwerte auf der Ebene der "Fach-Prüfer-Kombinationen": Eine "Fach-Prüfer-Kombination" wird gebildet durch die Prüflinge, die in einem bestimmten Prüfungsfach von demselben Prüfer geprüft werden. Eine "Fach-Prüfer-Kombination" wird dann ausgewertet, wenn sie mindestens 10 Prüflinge enthält. Da jedoch die für die Auswertung nicht unwesentliche Bildung von mehreren "Fach-Prüfer-Kombinationen" in einem Prüfungsfach dadurch erweitert wird, werden im VD noch eine Kombination mit 7 Prüflingen und im HD noch zwei Kombinationen mit 5 bzw. 8 Prüflingen aufgenommen. Nicht alle einzelnen Prüfungen sind einer Kombination zuordbar, da manche Prüfer nur ganz wenige Prüfungen in einem

Fach abnehmen. Die folgende Tabelle: E II. - 2 gibt die einzelnen "Fach-Prüfer-Kombinationen" in den Prüfungsfächern der beiden Prüfungsarten (VD und HD) wieder.

Tabelle: E II.- 2

"Fach-Prüfer-Kombinationen"

	Fach		Prüfer	Anzahl von Prüflingen
Vordiplom	Ausdr.	/	09	18
	Method.	/	02	16
	Allg.Psychol.	/	12	33
			14	7
	Entw.Psychol.	/	05	42
	Pers.Psychol.	/	02	12
			05	30
	Physiol.	/	08	40
	Philos.	/	01	10
			03	21
	Humang.	/	07	29
Hauptdiplom	Diagnostik	/	06	26
	Sozialpsych.	/	05	15
			12	13
	Päd.Psychol.	/	02	14
			09	14
	Klin.Psychol.	/	06	11
			18	15
	Angew.Psychol.	/	02	5
			04	8
			14	15

Die Prüfer sind in der Tabelle E II. - 2 durch Zahlen verschlüsselt. Wird im folgenden Text auf eine "Fach-Prüfer-Kombination" Bezug genommen, geschieht das durch die Angabe des Faches und des Prüferschlüssels: z.B. Pers.Psychol./05. Zwischen APO und NPO wird nicht unterschieden, wenn sich nur die Bezeichnung des

Faches ändert. In diesem Fall tritt das Fach unter der Bezeichnung der NPO auf; also Charakterkunde als Persönlichkeitspsychologie und Tiefenpsychologie als Klinische Psychologie.

2. Kennwerte auf der Ebene der Einzelfächer: Im VD sind es 7 Fächer und im HD 5 Fächer.

3. Kennwerte auf der Ebene aller Beurteilungen der Einzelfächer (im folgenden kurz als "Über-alle-Fächer" bezeichnet). Die Beurteilungen, die ein Prüfling für die verschiedenen Fächer abgibt, werden als voneinander unabhängige Elemente der Erhebung betrachtet. Die Zahl der Elemente, die jetzt nicht mehr Personen, sondern Prüfungen darstellen, steigt damit auf etwas weniger als das Siebenfache der Zahl der Prüflinge im VD (die Verringerung geht auf die Anerkennung von Nebenfächern, die die Prüflinge schon vorher absolviert haben, zurück) und auf das Fünffache der Zahl der Prüflinge im HD.

Durch die Bildung von Mittelwerten und ihre Verrechnung als neue Ausgangsdaten können zusätzliche Ebenen der Auswertung gewonnen werden.

4. Kennwerte auf der Ebene von Mittelwerten, die über alle Prüflinge einer "Fach-Prüfer-Kombination" hinweg berechnet werden.

5. Kennwerte auf der Ebene von Mittelwerten, die über alle Prüflinge eines Prüfungsfachs hinweg berechnet werden.

Diese (nicht erschöpfende) Aufstellung von Möglichkeiten läßt eine Entscheidung über sinnvolle Auswertungsebenen erst zu, wenn die Einflußquellen des Antwortverhaltens, die in die Kennwerte auf den einzelnen Ebenen unterschiedlich eingehen, berücksichtigt werden. Dabei lassen sich zwei Einflußquellen unterscheiden, die in Tabelle: E II. - 1 in Spalte I und II aufgeführt werden. Die Bedingungen 1. Art in Spalte I

sind Einflußquellen, die auf die einzelnen Daten wirken und die Unterschiedlichkeit (Varianz) der Daten zur Folge haben. Diese Bedingungen 1. Art sind eingebettet in Bedingungen 2. Art (Spalte II), die Rahmenbedingungen darstellen, indem sie für eine ganze Gruppe von Daten gleich sind. An einem Beispiel ist das zu erläutern: Für eine einzelne "Fach-Prüfer-Kombination" sind Bedingungen 1. Art die Prüflinge und die einzelnen Prüfungssituationen, die zu unterschiedlichen Beurteilungen der jeweiligen Einzelprüfung führen. Rahmenbedingung, also Bedingung 2. Art, ist z.B. der Prüfer, der alle Prüfungen dieser Kombination abnimmt. Einflußquellen als Bedingungen 1. Art werden durch eine Auswertung "innerhalb" der Daten einer Ebene deutlich; Einflußquellen als Bedingungen 2. Art werden durch die Auswertung "zwischen" Kennwerten deutlich.

Auf konkrete Auswertungsmethoden bezogen heißt das gewöhnlich: Bedingungen 1. Art werden durch Kontingenzen (Korrelationen) zwischen den Daten erfaßt; Bedingungen 2. Art werden als Unterschiede von Kontingenzen oder von Mittelwerten zwischen den einzelnen Bedingungen erfaßt.

Die Differenzierung ist notwendig, um in einem nächsten Schritt zu verdeutlichen, daß die Differenzierung von Einflußquellen auf Grund der empirischen Verklammerung der Quellen nur unzureichend möglich ist. Die Konfundierung der Quellen soll für jede der Ebenen analysiert werden. Allgemein geht sie auf drei verschiedenartige Konfundierungen zurück:

1. Die Prüflinge werden den einzelnen Prüfern nicht zufällig zugeordnet, sondern können die Prüfer (teilweise) wählen.

2. Die Prüfer werden den Fächern nicht zufällig zugeordnet, sondern bestimmte Prüfer prüfen nur bestimmte Fächer.

3. Die Prüfer werden den Prüfungsarten nicht

zufällig zugeordnet, sondern bestimmte Prüfer prüfen nur in einer der beiden Prüfungsarten.

Daraus ergeben sich die folgenden Probleme, die für die Auswertung auf den einzelnen Ebenen spezifisch sind.

Ebene "Fach-Prüfer-Kombination": Als Bedingungen 1. Art werden Unterschiede erfaßt, die auf Prüflinge oder Situationen (Prüfungen) und deren Interaktion zurückgehen. Durch die nicht zufällige Zuordnung der Prüflinge zu den "Fach-Prüfer-Kombinationen" entstehen zwei Komplikationen. Der Personeneinfluß als Bedingung 1. Art kann durch aktive Selektion zunehmen. Die Auswirkungen von Bedingungen 2. Art können in Wirklichkeit durch die aktive Selektion der Prüflinge vermittelt sein (Unterschiede zwischen Prüfern können auf Unterschiede der Studenten, die die Prüfer gewählt haben, zurückgehen). Der Einfluß der Prüferwahl ist allerdings dadurch begrenzt, daß nur in der Hälfte aller Einzelprüfungen eine solche Wahl möglich ist. Bei den Bedingungen 2. Art werden dagegen noch die beiden anderen Konfundierungen wirksam: Fach und Prüfer wie auch Prüfungsart und Prüfer lassen sich kaum klar voneinander trennen. Eine Trennung von Fach und Prüfer ist nur ansatzweise dort möglich, wo derselbe Prüfer in unterschiedlichen Fächern und unterschiedliche Prüfer in demselben Fach prüfen (s. Tabelle: E II. - 2).

Ebene "Einzelfach": Für die Bedingung 1. Art besteht das Problem der nicht zufälligen Zuordnung der Prüflinge nicht mehr, da jeweils alle Prüflinge in einem Fach in die Auswertung einbezogen werden. Dafür entsteht ein neues: Auch Unterschiede zwischen den Prüfern können jetzt in die Kennwerte (Korrelationen) eingehen. Dabei ist der mögliche Einfluß von Fach zu Fach sehr unterschiedlich je nachdem, ob im wesentlichen ein Prüfer alle Prüfungen abnimmt, oder ob mehrere Prüfer an den Prüfungen in einem Fach beteiligt sind. Auch bei

den Bedingungen 2. Art werden die Probleme nicht etwa geringer. Unterschiede zwischen den Prüfungsfächern können nicht einfach auf die strukturellen Bedingungen der Fächer zurückgeführt werden, weil gleichzeitig die Prüfung in bestimmten Prüfungsfächern an ganz bestimmte Prüfer gebunden ist. Die bei den "Fach-Prüfer-Kombinationen" ansatzweise mögliche Kontrolle der unterschiedlichen Bedeutung der Bedingung Prüfer und der Bedingung Prüfungsfach ist hier nicht mehr vorhanden.

Ebene "Über-alle-Fächer": Auf dieser Ebene werden die einzelnen Prüfungen als eigentliche Merkmalsträger (auch im Hinblick auf die Antworten der Prüflinge) betrachtet. Individuum/Situation, Prüfer und Prüfungsfach gehen ungeschieden als Bedingung 1. Art ein; dabei sind auch mögliche Interaktionen zwischen diesen Bedingungen eingeschlossen (was auch für die anderen Ebenen gilt und meist nicht gesondert erörtert wird). Probleme bestehen darin, daß die Bedingungen prinzipiell nicht mehr auseinandergehalten werden können. Als einzige Bedingung 2. Art fungiert noch die Prüfungsart, die aber weiterhin mit den Prüfern konfundiert ist.

Die Problematik der auf den beiden letzten Ebenen verwendeten Mittelwerte geht unmittelbar aus den bisherigen Erörterungen hervor: Als Bedingung 1. Art lassen sich Prüfer und Prüfungsfach nicht trennen. Bei den Mittelwerten für die "Fach-Prüfer-Kombinationen" kann zusätzlich die Auswahl des Prüfers durch die Prüflinge hineinspielen.

Insgesamt gesehen wird deutlich, daß keine Ebene eine problemlose Erfassung der Bedingungen ermöglicht. Ziel der Untersuchung der Einzelprüfungen muß es sein, im Hinblick auf die Wirkung der Noten die individuell-situative Komponente von den Einflußquellen Fach, Prüfer und Prüfungsart zu trennen. Das ge-

lingt auf der Ebene der "Fach-Prüfer-Kombinationen" besser als auf der Ebene "Einzelfach". Für die Wahl der Ebene "Fach-Prüfer-Kombination" spricht auch:

1. daß wenigstens ansatzweise eine Abwägung der Bedingung Prüfer und Prüfungsfach erfolgen kann;
2. daß sich bei einer Untersuchung der Interaktionen in mündlichen Prüfungen sehr deutliche Unterschiede zwischen den Prüfern ergeben haben (SCHLIEHE & SEEGER 1974);
3. daß, wie noch nachgewiesen wird, sich die Notengebung der einzelnen Prüfer deutlich unterscheidet.

Zusätzlich aber sollte durch eine Auswertung auf der Ebene "Über-alle-Fächer" das Ausmaß der in die gleiche Richtung weisenden Einflüsse aller Bedingungen ermittelt werden. Durch Vergleich der für die "Fach-Prüfer-Kombinationen" und der "Über-alle-Fächer" ermittelten Werte kann abgeschätzt werden, in welchem Verhältnis individuell-situative Einflüsse und übergreifende situative Einflüsse (Fach, Prüfer) stehen.

Wenn zusätzlich eine Auswertung der Mittelwerte als Daten erfolgen soll, scheint aus zwei Gründen eine Verwendung der "Mittelwerte der Fach-Prüfer-Kombinationen" angeraten:

1. Sie lassen sich mit den Ergebnissen der Auswertung in den "Fach-Prüfer-Kombinationen" verknüpfen.
2. Die Zahl der Prüfungsfächer ist zu gering für Korrelationen, die als Auswertungsmethode für die Mittelwerte in Frage kommen.

Die Festlegung der drei Auswertungsebenen: "Fach-Prüfer-Kombinationen", "Über-alle-Fächer" und "Mittelwerte der Fach-Prüfer-Kombinationen" ist zusätzlich mit inferenzstatistischen Problemen befrachtet. Die Daten, die auf der Ebene "Über-alle-Fächer" und

"Mittelwerte der Fach-Prüfer-Kombinationen" inferenzstatistisch bearbeitet werden, erfüllen die in die inferenzstatistischen Modelle eingehende Annahme von der Unabhängigkeit der einzelnen Messungen nicht. Beispielsweise ist in der Auswertung "Über-alle-Fächer", bei der jeder Prüfling im HD fünfmal und im VD fast siebenmal berücksichtigt wird, die einzelne Einschätzung einer Prüfung durch den Prüfling nicht von den anderen Einschätzungen desselben Prüflings unabhängig. Auch in die Mittelwerte, die für die einzelnen "Fach-Prüfer-Kombinationen" berechnet werden, gehen in teilweiser Überschneidung je nach Übereinstimmung der Prüflinge zwischen zwei "Fach-Prüfer-Kombinationen" dieselben Prüflinge ein. Insofern sind diese Mittelwerte gleichfalls als nicht voneinander unabhängig zu betrachten.

Die Auswirkungen der Verwendung teilweise abhängiger Werte sind je nach Auswertungsart und Grad der Abhängigkeit unterschiedlich. Bei Unterschieden zwischen Kennwerten bedingen sie eher eine konservative Testung, bei den Kennwerten selbst können sie auch in die andere Richtung wirken. Das Ausmaß der Verfälschung wird von dem empirischen Ausmaß der Abhängigkeit bestimmt, das in Teil B für die Noten und Teil E für die abhängigen Variablen angegeben wird (s. Kapitel III.I.).

Auf alle inferenzstatistischen Angaben, die solche partiell abhängigen Daten zur Grundlage haben, wird mit einem (?) aufmerksam gemacht.

Anhang I: Wortlaut der Fragen (PVF und PFF)

Die Fragen aus den beiden Fragebögen, die in dieser Arbeit ausgewertet werden, sind wörtlich wiedergegeben. Bei Fächerfragen (F 1, F 2 ..) ist nur jeweils die Antwortskala für das erste Prüfungsfach dargestellt.

Die Fragen sind unterteilt in die thematischen Fragen, die den abhängigen Variablen von Kapitel III.I. und III.II. entsprechen, und die übrigen Fragen, die in den Kapiteln III.II. und III.III. in die Auswertung einbezogen werden. Die durch Unterstreichung hervorgehobene Kurzbezeichnung entspricht den Vereinbarungen von Exkurs I. Sie ist nicht Teil des Fragebogens, sondern dient der Verständigung bei der Erwähnung der einzelnen Fragen im laufenden Text.

Die Reihenfolge der Fragen ist die Reihenfolge im Fragebogen. Doch sind in vielen Fällen den dargestellten Fragen noch weitere zwischengeschaltet, die in dieser Auswertung nicht berücksichtigt werden.

A I.1.: Thematische Fragen

Prüfungsvorbereitungsfragebogen

PVF-MOT.BEL. (Motivation: Beliebtheit)

Wie gern haben Sie sich während des Studiums mit den einzelnen Fächern beschäftigt?

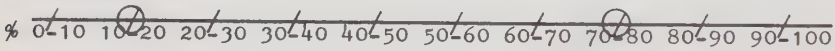
VD	HD	sehr gern	gern	wenig gern	nicht gern
Met/A	Dia	o	o	o	o

(Die Kurzbezeichnungen für das erste Prüfungsfach bedeuten: im Vordiplom "Met" das Fach Methodenlehre der neuen Prüfungsordnung, "A" das Fach Ausdruckskunde der alten Prüfungsordnung; im Hauptdiplom "Dia" das Fach Psychologische Diagnostik. Wie angemerkt, sind die übrigen Prüfungsfächer nicht mehr einzeln aufgeführt.)

PVF-SK.FÄCHER (fachbezogenes Selbstkonzept eigener Fähigkeiten)

Wir möchten feststellen, wie gut Studenten in der Lage sind, ihre eigenen Fähigkeiten in den verschiedenen Gebieten einzuschätzen. Besinnen Sie sich bitte auf Ihr bisheriges Studium, die Prüfungsvorbereitung usw. Tragen Sie Ihre Selbsteinschätzung in eine Prozentrangskala durch Umkreisen einer der bezeichneten Stellen ein.

Beispiel:



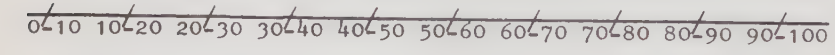
Ein Kreis an der Stelle, die mit 10-20 gekennzeichnet ist, bedeutet, daß Sie sich in diesem Gebiet zu den untersten 20 % gehörig fühlen; ein Kreis an der Stelle 70-80 bedeutet, daß Sie sich zu den obersten 30 % rechnen.

VD HD
Met/A Dia

PVF-SK.GESAMT (allgemeines Selbstkonzept eigener Fähigkeiten)

(Anweisung: s. PVF-SK.FÄCHER)

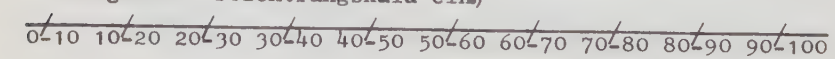
Psychologie insgesamt



PVF-SK.ZUKUNFT (zukunftsbezogenes Selbstkonzept)

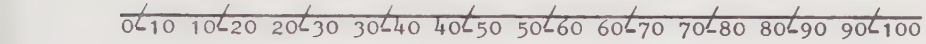
Vordiplom

Mit welchen Studienleistungen rechnen Sie im Vergleich mit Ihren Mitstudenten nach dem VD? (Kreisen Sie die Stelle, die Ihre Erwartungen wiedergibt, auf der folgenden Prozentrangskala ein.)



Hauptdiplom

Versuchen Sie einmal, sich vorzustellen, welche berufliche Stellung Sie in 10 Jahren erreicht haben werden. Bezeichnen Sie die Höhe der Stellung in der folgenden Prozentrangskala. Vergleichsgruppe sollen dabei alle beruflich tätigen Psychologen sein.



PVF-MOT.W.A. (Motivation: Weiterarbeit)

Mit welcher Intensität wollen Sie nach der Prüfung in den einzelnen Fächern weiterarbeiten? (Gemeint sind die Fächer, in denen Sie sich jetzt prüfen lassen.)

VD	HD	sehr intensiv	intensiv	wenig intensiv	nicht intensiv
Met/A	Dia	o	o	o	o

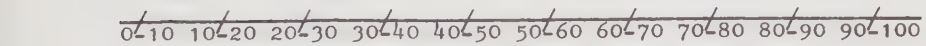
Prüfungsfolgenfragebogen

PFF-SK.FÄCHER (fachbezogenes Selbstkonzept eigener Fähigkeiten)

Wir möchten feststellen, wie gut Studenten in der Lage sind, ihre eigenen Fähigkeiten in den verschiedenen Gebieten einzuschätzen. Besinnen Sie sich bitte auf die Prüfung, Ihr bisheriges Studium usw. Tragen Sie Ihre Selbsteinschätzung in eine Prozentrangskala durch Umkreisen einer der bezeichneten Stellen ein.

(Anweisung: s. PVF-SK.FÄCHER)

VD HD
Met/A Dia



PFF-SK.GESAMT (allgemeines Selbstkonzept eigener Fähigkeiten)

(Anweisung: s. PVF-SK.FÄCHER)

Psychologie insgesamt

0-10 10-20 20-30 30-40 40-50 50-60 60-70 70-80 80-90 90-100

PFF-MOT.W.A. (Motivation: Weiterarbeit)

Mit welcher Intensität werden Sie in den einzelnen Fächern weiterarbeiten? (Gemeint sind die Fächer, in denen Sie sich jetzt haben prüfen lassen.)

VD	HD		sehr intensiv	intensiv	wenig intensiv	nicht intensiv
Met/A	Dia	....	o--------o	--------o	--------o	--------o

PFF-SK.ZUKUNFT (zukunftsbezogenes Selbstkonzept)

(identisch mit PVF-SK.ZUKUNFT)

PFF-MOT.BEL. (Motivation: Beliebtheit)

Wie gern würden Sie sich im Augenblick mit den einzelnen Fächern beschäftigen? (Bitte geben Sie Ihr Urteil als VD-Absolvent für die Fächer des VD, als HD-Absolvent für die Fächer des HD ab.)

VD	HD		sehr gern	gern	wenig gern	nicht gern
Met/A	Dia	.....	o--------o	--------o	--------o	--------o

A I.2.: Übrige Fragen

Die auf die einzelnen Prüfungsfächer bezogenen Fragen und die auf die Gesamtprüfung bezogenen Fragen sind, weil dies die Darstellung erleichtert, nicht noch einmal getrennt. Ihre Kurzbezeichnung aber ist unterschiedlich.

Prüfungsvorbereitungsfragebogen

PVF - F 1. ATMOSPHÄRE

Schätzen Sie bitte das erwartete Verhalten Ihrer Prüfer in den einzelnen Fächern ein

		Atmosphäre der Prüfung	
VD	HD	gespannt	entspannt
Met/A	Dia	o--------o--------o--------o	

PVF - F 2. ANFORDERUNGEN

(Anweisung: s. PVF - F 1.)

		Anforderungen des Prüfers	
VD	HD	hoch	niedrig
Met/A	Dia	o--------o--------o--------o	

PVF - F 3. FESTLEGUNG

In welchem Ausmaß halten Sie die Auswahl des Prüfungsstoffs in den einzelnen Fächern für festgelegt?

		sehr viel	viel	etwas	nicht
VD	HD				
Met/A	Dia	o--------o--------o--------o			

PVF - F 4. ANNEHMLICHKEIT

Wie angenehm war (ist) die Vorbereitung in den einzelnen Fächern?

VD	HD	angenehm	unangenehm
Met/A	Dia	o--------o--------o--------o	

PVF - F 5. RELEVANZ

Sie müssen sich für die Prüfung ganz bestimmte Wissensinhalte aneignen. Welchen Anteil haben daran jene Inhalte, die Sie allein für die Prüfung lernen

müssen, ohne später etwas damit anfangen zu können?

(Hinweis: Wenn Sie 0-25 % ankreuzen, heißt das, daß 0-25 % allein für die Prüfung gelernt wurden.)

VD	HD	0 - 25 %	25 - 50 %	50 - 75 %	75 - 100 %
Met/A	Dia	o--------o	--------o	--------o	--------o

PVF - F 6. WICHTIGKEIT

Wie wichtig war für Sie die Vorbereitung in den einzelnen Fächern im Hinblick auf Beruf bzw. weiteres Studium?

VD	HD	sehr wichtig	wichtig	wenig wichtig	nicht wichtig
Met/A	Dia	o--------o	--------o	--------o	--------o

PVF - F 7. PRÜFUNGSANGST

Beurteilen Sie bitte die Fächer danach, wieviel Prüfungsangst sie jetzt bei Ihnen erzeugen

VD	HD	sehr viel	viel	einige	wenig- keine
Met/A	Dia	o--------o	--------o	--------o	--------o

PVF - F 8. NOTENERWARTUNG

Tragen Sie bitte Ihre Notenerwartungen für die Einzelprüfungen und die Gesamtprüfung ein. Versuchen Sie, den Zweckpessimismus eines jeden Prüflings so weit wie möglich zurückzustellen und mehr Ihre geheimen Vorstellungen wiederzugeben.
(Unterstreichen Sie bitte die jeweils erwartete Note.)

VD	HD					
Met/A	Dia	s.gut	gut	befr.	ausr.	n.ausr.

PVF - I, NOTENERW.GESAMT (Notenerwartung für die Gesamtprüfung)

(Anweisung: s. PVF - F 8.)

Gesamtprüfung	s.gut	gut	befr.	ausr.	n.ausr.

PVF - II, NOTENSKALA

Nehmen Sie einmal an, vor einer mündlichen Prüfung gibt Ihnen der Prüfer die Gelegenheit, zwischen zwei verschiedenen Prüfungsbedingungen zu wählen: einer Bedingung A und einer Bedingung B. In beiden Bedingungen bekommen Sie mit gleicher Wahrscheinlichkeit eine von zwei Noten.

Beispiel:

A	B
1	3
4	3

In der Bedingung A können Sie mit 50 % Wahrscheinlichkeit mit der Note 1 (sehr gut) und mit 50 % Wahrscheinlichkeit mit der Note 4 (ausreichend) rechnen (die Noten 1 und 4 sind damit gleichwahrscheinlich).
In der Bedingung B können Sie jedesmal mit einer Wahrscheinlichkeit von 50 % mit einer 3 (befriedigend) rechnen.

Im folgenden finden Sie 15 solcher fiktiven Bedingungspaare. Kreuzen Sie bitte bei jedem Bedingungspaar den Buchstaben der Bedingung an, die Sie der jeweils anderen vorziehen würden. Ziehen Sie keine der beiden Bedingungen vor, dann kreuzen Sie bitte beide Buchstaben an.

A	B	A	B	A	B	A	B	A	B
1	2	2	3	4	3	3	2	2	1
3	2	4	3	4	5	4	5	4	5
A	**B**	**A**	**B**	**A**	**B**	**A**	**B**	**A**	**B**
4	1	1	2	1	2	2	1	3	2
4	5	4	2	4	3	2	5	3	5
A	**B**	**A**	**B**	**A**	**B**	**A**	**B**	**A**	**B**
2	1	1	3	4	2	3	1	3	1
3	5	4	3	4	5	4	5	3	5

Prüfungsfolgenfragebogen

PFF - III. NOTENSKALA

(identisch mit PVF - II. NOTENSKALA; s. oben)

PFF - IV. LEISTUNG.GESAMT

In welchem Verhältnis zu dem, was Sie wirklich können, steht die Gesamtnote, die Sie in der Prüfung erhalten haben? (Geben Sie Ihr Urteil bitte auf der folgenden Skala durch Ankreuzen einer der bezeichneten Stellen wieder.)

eher zu gut	zu- treffend	eher zu schlecht
o--------o--------o--------o--------o		

PFF - F 9. WICHTIGKEIT

Wie wichtig war für Sie die Vorbereitung in den einzelnen Fächern im Hinblick auf Beruf bzw. späteres Studium?

VD	HD	sehr wichtig	wichtig	wenig wichtig	nicht wichtig
Met/A	Dia	o--------o--------o--------o			

(Die folgenden Fragen werden im PFF zusammenhängend jeweils für die einzelnen Prüfungsfächer gestellt.)

Fach:	VD	HD
	Met/A	Dia

PFF - F 10. WIRKL.LEISTUNG

Verglichen mit meinen wirklichen Leistungen in der Prüfung ist die Bewertung meiner Prüfungsleistungen

zu gut	eher zu gut	zu- treffend	eher zu schlecht	zu schlecht
o--------o--------o--------o--------o				

PFF - F 11, STUDIENLEISTUNG

Verglichen mit meinen bisherigen Studienleistungen ist die Bewertung meiner Prüfungsleistungen

zu gut	eher zu gut	zu-treffend	eher zu schlecht	zu schlecht
o	o	o	o	o

PFF - F 12, RÜCKM,NEG.

Die Prüfung hat mich auf Wissenslücken und Verständnisschwierigkeiten in dem Fach aufmerksam gemacht.

trifft voll zu	trifft weitgehend zu	trifft etwas zu	trifft nicht zu
o	o	o	o

PFF - F 13, RÜCKM.POS,

Die Prüfung hat mich auf meine besonderen Stärken in Kenntnissen und Verständnis aufmerksam gemacht.

trifft voll zu	trifft weitgehend zu	trifft etwas zu	trifft nicht zu
o	o	o	o

PFF - F 14, KONGRUENZ,VORB.

Das, was in der Prüfung verlangt wurde, entsprach dem, womit ich mich während der Prüfungsvorbereitung beschäftigt hatte.

trifft voll zu	trifft weitgehend zu	trifft etwas zu	trifft nicht zu
o	o	o	o

PFF - F 15, KONGRUENZ,STUD.

Das, was in der Prüfung an Inhalten verlangt wurde, entspricht dem, was in einem sinnvollen Studium auch vermittelt werden sollte.

trifft voll zu | trifft weitgehend zu | trifft etwas zu | trifft nicht zu

o----------o----------o----------o

PFF - F 16. ANPASSUNG

Haben Sie in der Prüfung Ihre eigene Meinung vertreten, oder haben Sie sich der erwarteten Meinung des Prüfers angepaßt?

eigene Meinung voll vertreten — ganz der Meinung d.Prüfers angepaßt

o----------o----------o----------o

PFF - F 17. ATMOSPHÄRE

Wie hat sich Ihr Prüfer verhalten?

Atmosphäre der Prüfung

gespannt — entspannt

o--------o--------o--------o

PFF - F 18. ANFORDERUNGEN

Wie hat sich Ihr Prüfer verhalten?

Anforderungen des Prüfers

hoch — niedrig

o--------o--------o--------o

PFF - F 19. PRÜF.INSGESAMT

Wie war die Prüfung insgesamt?

angenehm — unangenehm

o--------o--------o--------o

Literaturverzeichnis

ABERCROMBIE, J. Teaching, examining, and the curriculum. Universities Quarterly, 1964/65, 19, 143-149.

AGER, M., & WELTMAN, J. The present structure of university examinations. Universities Quarterly, 1966/67, 21, 272-285.

AHMANN, J. S., & GLOCK, M. D. Evaluating pupil growth: Principles of tests and measurements. Boston: Allyn & Bacon, 1967.

AIKEN, L. R. Jr. The grading behavior of a college faculty. Educational and Psychological Measurement, 1963, 23, 319-322.

ALBERT, H. Modell-Platonismus. Der neoklassische Stil des ökonomischen Denkens in kritischer Beleuchtung. In E. Topitsch (Ed.), Logik der Sozialwissenschaften. Köln: Kiepenheuer & Witsch, 1965.

ALEXANDER, E. D. The marking system and poor achievement. In R. K. Parker (Ed.), Readings in educational psychology. Boston: Allyn & Bacon, 1968.

ALVES, H., & LODDERS, J. Inhaltliche Studiensteuerung: Eine Untersuchung über die Steuerungswirkung von Prüfungen. Unveröffentlichte Diplomarbeit, Psychologisches Institut der Universität Münster, 1974.

AMATORA, M. Validity in self-evaluation. Educational and Psychological Measurement, 1956, 16, 119-126.

AMELANG, M. Ausbildung und Prüfung im Fach Psychologie - Versuch einer Bewertung, angestellt am Beispiel eines Instituts. Psychologische Rundschau, 1973, 24, 169-182.

AMMONS, R. B. Effects of knowledge of performance: A survey and tentative theoretical formulation. Journal of General Psychology, 1956, 54, 279-299.

ANGER, H. Befragung und Erhebung. In C. F. Graumann (Ed.), Handbuch der Psychologie. Bd. 7.1. Sozialpsychologie. Göttingen: Hogrefe, 1969.

APPELBAUM, S. A. Science and persuasion in the psychological test report. Journal of Consulting and Clinical Psychology, 1970, 35, 349-355.

ARONSON, E., & CARLSMITH, J. M. Performance expectancy as a determinant of actual performance. Journal of Abnormal and Social Psychology, 1962, 65, 178-182.

ARSENIAN, S. Own estimate and objective measurement. Journal of Educational Psychology, 1942, 33, 291-302.

ASCHERSLEBEN, K. Zur Statistik der Schulnotengebung. Die Deutsche Schule, 1970, 62, 761-768.

ATHANASIADIS, T. Zum System der Leistungsbewertung mit Zensuren. Wiesbaden: Steiner, 1974.

ATKINSON, J. W. An introduction to motivation. Princeton: Van Nostrand, 1964.

ATKINSON, J. W., & FEATHER, N. T. (Eds.) A theory of achievement motivation. New York: Wiley, 1966.

BADER, P. Psychologisches und Pädagogisches über Zensuren. Zeitschrift für pädagogische Psychologie und experimentelle Pädagogik, 1913, 14, 380-396.

BAHRDT, H. P. Leistungsstand, Leistungsdefizit und Leistungskontrolle bei den Studenten von heute. Deutsche Universitätszeitung/Hochschul-Dienst. 1972, [27], 536-542.

BAILEY, R. C., & BAILEY, K. G. Perceived ability in male and female college students. Perceptual and Motor Skills, 1971, 32, 293-294.

BAILEY, K. G., & GIBBY, R. G. Developmental differences in self-ratings on intelligence. Journal of Clinical Psychology, 1971, 27, 51-54.

BAIRD, L. L. The relation of vocational interests to life goals, self-ratings of ability and personality traits, and potential for achievement. Journal of Educational Measurement, 1970, 7, 233-239.

BAK (Bundesassistentenkonferenz) (Ed.) Forschendes Lernen - Wissenschaftliches Prüfen. Bonn: BAK, 1970.

BAKER, R. L., & DOYLE, R. Elementary school marking practices and subsequent high school achievement. Educational Administration and Supervision, 1958, 44, 158-166.

BALCH, J. The influence of the evaluation instrument on students' learning. American Educational Research Journal, 1964, 1, 169-182.

BARNES, E. J., & PRESSEY, S. L. The reliability and validity of oral examinations. School and Society, 1929, 30, 719-722.

BARTEL, H. Statistik II für Psychologen, Pädagogen und Sozialwissenschaftler. Stuttgart: Fischer, 1972.

BASS, B. M. Intrauniversity variations in grading practices. Journal of Educational Psychology, 1951, 42, 366-368.

BEARD, R. M. Research into teaching methods in higher education. London: Society for Research in Higher Education, 1968.

BEARD, R.[M.] Lehren und Lernen an der Hochschule. Düsseldorf: Schwann, 1972.

BEARD, R. M., LEVY, P. M., & MADDOX, H. Academic performance at university. Educational Review, 1964, 16, 163-174.

BECKER, S. W., & SIEGEL, S. Utility of grades: Level of aspiration in a decision theory context. Journal of Experimental Psychology, 1958, 55, 81-85.

BECKER, S. W., & SIEGEL, S. Utility and level of aspiration. American Journal of Psychology, 1962, 75, 115-120.

BEHREND, F. (Ed.) Vom Sinn und Unsinn des Berechtigungswesens. Leipzig: Quelle & Meyer, 1929.

BENDIG, A. W. The reliability of self-ratings as a function of amount of verbal anchoring and of the number of categories on the scale. Journal of Applied Psychology, 1953, 37, 38-41.

BENDIG, A. W. Comparative validity of objective and projective measures of need achievement in predicting students' achievement in introductory psychology. Journal of General Psychology, 1959, 60, 237-243.

BENJAMINS, J. Changes in performance in relation to influences upon self-conceptualization. Journal of Abnormal and Social Psychology, 1950, 45, 473-480.

BERDIE, R. F. Changes in self-ratings as a method of evaluating counseling. Journal of Counseling Psychology, 1954, 1, 49-54.

BERGIN, A. E. The effect of dissonant persuasive communication upon changes in self-referring attitudes. Journal of Personality, 1962, 30, 423-438.

BERGIUS, R. (Ed.) Handbuch der Psychologie. Bd. 1.2. Lernen und Denken. Göttingen: Hogrefe, 1964.

BERLYNE, D. E. Konflikt, Erregung, Neugier. Stuttgart: Klett, 1974.

BETZ, D. Rhythmische Schwankungen als Fehler in der Notengebung bei mündlichen Prüfungen. Psychologie in Erziehung und Unterricht, 1974, 21, 1-14.

BEYER, H. N. Effect of students' knowledge of their predicted grade point averages on academic achievement. Journal of Counseling Psychology, 1971, 18, 603-605.

BIÄSCH, H. Psychologische Hintergründe des Examensverhaltens. Jahrbuch für Psychologie und Psychotherapie, 1968, 16, 115-129.

BIGGS, D. A., & JOHNSON, J. Self-made academic predictions of junior college students. Journal of Educational Research, 1972, 66, 85-88.

BIGGS, D. A., & TINSLEY, D. J. Student-made academic predictions. Journal of Educational Research, 1970, 63, 195-197.

BILBY, R. W., BROOKOVER, W. B., & ERICKSON, E. L. Characterizations of self and student decision making. Review of Educational Research, 1972, 42, 505-524.

BIRNEY, R. C. The effects of grades on students. Journal of Higher Education, 1964, 35, 96-98.

BIXLER, H. H. School marks. Review of Educational Research, 1936, 6, 169-173 und 247-248.

BLACK, P. J. University examinations. Physics Education, 1968, 3, 93-99.

BLANKERTZ, H. Theorien und Modelle der Didaktik. München: Juventa, 1969.

BLOOM, B. S. (Ed.) Taxonomie von Lernzielen im kognitiven Bereich. Weinheim: Beltz, 1972.

BLOOM, B. S., HASTINGS, J. T., & MADAUS, G. F. (Eds.) Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.

BOKELMANN, H. Maßstäbe pädagogischen Handelns. Würzburg: Werkbund-Verlag, 1965.

BORISLOW, B. Self-evaluation and academic achievement. Journal of Counseling Psychology, 1962, 9, 246-254.

BOSHIER, R. The effect of academic failure on self-concept and maladjustment indices. Journal of Educational Research, 1972, 65, 347-351.

BOURDIEU, P., & PASSERON, J.-C. Die Illusion der Chancengleichheit. Stuttgart: Klett, 1971.

BREHM, J. W., & COHEN, A. R. Explorations in cognitive dissonance. New York: Wiley, 1962.

BRIM, O. G. Jr. College grades and self-estimates of intelligence. Journal of Educational Psychology, 1954, 45, 477-484.

BRIM, O. G. Jr. High and low self-estimates of intelligence. In O. G. Brim Jr., R. S. Crutchfield, & W. H. Holtzman, Intelligence: Perspectives 1965. New York: Harcourt, Brace & World, 1966.

BRIM, O. G. Jr. American attitudes toward intelligence tests. In J. T. Flynn & H. Garber (Eds.), Assessing behavior. Readings in educational and psychological measurement. Reading: Addison-Wesley, 1967.

BRIM, O. G. Jr., NEULINGER, J., & GLASS, D. C. Experiences and attitudes of american adults concerning standardized intelligence tests. New York: Russell Sage Foundation, 1965.

BROOKOVER, W. B., & THOMAS, S. Self-concept of ability and school achievement. Sociology of Education, 1964, 37, 271-278.

BROWN, C. H., & GELDER, D. V. Emotional reactions before examinations. Journal of Psychology, 1938, 5, 1-29.

BROWN, F. G. Measurement and evaluation. Itasca: Peacock, 1971.

BULL, G. M. An examination of the final examination in medicine. The Lancet, 1956, 271, 368-372.

BUNGARD, W., & LÜCK, H. E. Forschungsartefakte und nicht-reaktive Meßverfahren. Stuttgart: Teubner, 1974.

BURGER, R. Begabtenfindung und Benotungspraxis. Schule und Psychologie, 1967, 14, 329-344.

CAPPEL, W. Untersuchungen zur Validität der 1. Lehrerprüfung. Kaiserslautern: Pädagogische Hochschule, Reden, Heft 2, 1967.

CARTER, H. D. How reliable are good oral examinations? Californian Journal of Educational Research, 1962, 13, 147-153.

CASHEN, V. M. Students', parents', and counselors' prediction of academic success. Journal of Educational Research, 1967, 60, 212-224.

CATRON, D. W. Educational-vocational group counseling: The effects on perception of self and others. Journal of Counseling Psychology, 1966, 13, 202-207.

CHAIKIN, A. L. The effects of four outcome schedules on persistence, liking for the task, and attribution of causality. Journal of Personality, 1971, 39, 512-526.

CHANSKY, N. M. Perception of grade intervals. Psychological Reports, 1969, 25, 566.

CHARVAT, J., McGUIRE, C. H., & PARSONS, V. A review of the nature and uses of examinations in medical education. Geneva: WHO, 1968.

CLARK, D. C. Competition for grades and graduate-student performance. Journal of Educational Research, 1969, 62, 351-354.

CLARK, P. E. Can college students grade themselves? School and Society, 1938, 47, 614-616.

CLAUSS, G., & EBNER, H. Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen. Frankfurt: Deutsch, 1975.

CLEMENS-LODDE, B. Einstellungen von Lehrenden zu intrinsisch und extrinsisch motivierendem Lehrverhalten. Dissertation, Münster, 1974.

COFER, C. N., & APPLEY, M. H. Motivation: Theory and research. New York: Wiley, 1964.

COLEMAN, J. S. Academic achievement and the structure of competition. Harvard Educational Review, 1959, 29, 330-351.

COMBS, A. W., SOPER, D. W., & COURSON, C. C. The measurement of self concept and self report. Educational and Psychological Measurement, 1963, 23, 493-500.

COMPTON, R. K. Student evaluation of knowing college aptitude test score. Journal of Educational Psychology, 1941, 32, 656-664.

COOMBS, R. H., & DAVIES, V. Self-conception and the relationship between high school and college scholastic achievement. Sociology and Social Research, 1966, 50, 460-471.

COX, R. Resistance to change in examining. Universities Quarterly 1966/67, 21, 352-358. (a)

COX, R. Examinations and higher education: A survey of the literature. Universities Quarterly, 1966/67, 21, 292-340. (b)

CRONBACH, L. J., & FURBY, L. How we should measure "change"- or should we? Psychological Bulletin, 1970, 74, 68-80.

CRONBACH, L. J., & GLESER, G. C. Psychological tests and personnel decisions. Urbana: University of Illinois Press, 1957.

CROWNE, D. P., & MARLOWE, D. A. The approval motive: Studies in evaluative dependence. New York: Wiley, 1964.

CROWNE, D. P., & STEPHENS, M. W. Self-acceptance and self-evaluative behavior: A critique of methodology. Psychological Bulletin, 1961, 58, 104-121.

DÄUMLING, A. M. Psychologische Probleme des Hochschul-examens. Jahrbuch für Psychologie und Psychotherapie, 1968, 16, 104-115.

DANN, H.-D., & MÜLLER-FOHRBRODT, G. Einige Ursachen und Folgen der Benotungspraxis an Gymnasien. Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 1972, 4, 261-270.

DAVIDSON, H. H., & LANG, G. Children's perceptions of their teachers' feelings toward them related to self-perception, school achievement and behavior. Journal of Experimental Education, 1960, 29, 107-118.

DAVIS, F. B. Educational measurements and their interpretation. Belmont: Wadsworth, 1964.

DAVIS, J. A. Use of measurement in student planning and guidance. In R. L. Thorndike (Ed.), Educational measurement. Washington: American Council on Education, 1971.

DAVIS, W. L., & DAVIS, D. E. Internal-external control and attribution of responsibility for success and failure. Journal of Personality, 1972, 40, 123-136.

DECI, E. L. Intrinsic motivation, extrinsic reinforcement, and inequity. Journal of Personality and Social Psychology, 1972, 22, 113-120.

DEUTSCH, M. Social relations in the classroom and grading practices. Journal of Educational Research, 1951, 45, 145-152.

DEUTSCH, M., & SOLOMON, L. Reactions to evaluations by others as influenced by self-evaluations. Sociometry, 1959, 22, 93-112.

DILLER, L. Conscious and unconscious self-attitudes after success and failure. Journal of Personality, 1954, 23, 1-12.

DÖRNER, K. Die Hochschulpsychiatrie. Stuttgart: Enke, 1967.

DOHSE, W. Das Schulzeugnis. Weinheim: Beltz, 1967.

DORIS, J. Test anxiety and blame assignment in grade school children. Journal of Abnormal and Social Psychology, 1959, 58, 181-190.

DORIS, J., & SARASON, S. B. Test anxiety and blame assignment in a failure situation. Journal of Abnormal and Social Psychology, 1955, 50, 335-338.

DOUBRAWA, R. Zusammenhänge zwischen Abitur- und Examensnoten bei Psychologiestudenten. Psychologische Rundschau, 1971, 22, 79-97.

DREITZEL, H. P. Soziologische Reflexionen über das Elend des Leistungsprinzips. In Sinn und Unsinn des Leistungsprinzips. München: Deutscher Taschenbuch Verlag, 1974.

DRESSEL, P. L. Comprehensive examination programs. In P. L. Dressel (Ed.), Evaluation in higher education. Boston: Houghton Mifflin, 1961.

DUIJKER, H. C. J., & van RIJSWIJK, M. J. Trilingual psychological dictionary. Vol. 1. English/French/German. Bern: Huber, 1975.

DUMKE, D. Schülerleistung und Zensur. Hannover: Schroedel, 1974.

DUVAL, S., & WICKLUND, R. A. A theory of objective self awareness. New York: Academic Press, 1972.

EAGLY, A. H., & ACKSEN, B. A. The effect of expecting to be evaluated on change toward favorable and unfavorable information about oneself. Sociometry, 1971, 34, 411-422.

EBEL, R. L. Using exams to promote learning. In R. M. Cooper (Ed.), The two ends of the log. Minneapolis: University of Minneapolis Press, 1958.

ECKSTEIN, B. Hochschuldidaktik und gesamtgesellschaftliche Konflikte. Frankfurt: Suhrkamp, 1972.

EDDING, F. Die Hochschulen im Wachsen der Wirtschaft. Recht und Wirtschaft der Schule, 1962, 3.

EELLS, W. C. An improvement in the theoretical basis of five point grading systems based on the normal probability curve. Journal of Educational Psychology, 1930, 21, 128-135.

EIGLER, G. Das Verhältnis von Lernen und Leistung als Forschungsproblem. In H. Roth & D. Friedrich (Eds.), Bildungsforschung, Teil 1. Bd. 50, Gutachten und Studien der Bildungskommission, Deutscher Bildungsrat. Stuttgart: Klett, 1975.

ERBSLÖH, E. Techniken der Datensammlung 1. Interview. Stuttgart: Teubner, 1972.

ESKOLA, A. Level of aspiration in career selection. Acta Sociologica, 1961, 5, 239-248.

EVANS, L. R., INGERSOLL, R. W., & SMITH, E. J. The reliability, validity and taxonomic structure of the oral examination. Journal of Medical Education, 1966, 41, 651-657.

FALK, R. Zur Psychologie der schulischen Leistungsbeurteilung durch Zensierung. Wissenschaftliche Zeitschrift der Universität Halle-Wittenberg, Gesellschaftlich-Sprachwissenschaftliche Reihe, 1962, 11, 1015-1032.

FAY, P. J. The effect of the knowledge of marks on the subsequent achievement of college students. Journal of Educational Psychology, 1937, 38, 548-554.

FELDMAN, S. (Ed.) Cognitive consistency. Motivational antecedents and behavioral consequences. New York: Academic Press, 1966.

FEND, H. Konformität und Selbstbestimmung. Weinheim: Beltz, 1971.

FESTINGER, L. A theory of social comparison processes. Human Relations, 1954, 7, 117-140.

FESTINGER, L. A theory of cognitive dissonance. Stanford: Stanford University Press, 1957.

FESTINGER, L., TORREY, J., & WILLERMAN, B. Self-evaluation as a function of attraction to the group. Human Relations, 1954, 7, 161-174.

FISCH, R. Konfliktmotivation und Examen. Meisenheim: Hain, 1970.

FITCH, M. L., DRUCKER, A. J., & NORTON, J. A. Jr. Frequent testing as a motivating factor in large lecture classes. Journal of Educational Psychology, 1951, 42, 1-20.

FITTKAU, B., & LANGER, H. Auswirkungen schriftlicher Ermutigung unter Klassenarbeiten auf Angst und Leistung der Schüler. Psychologie in Erziehung und Unterricht, 1974, 11, 15-21.

FLAMMER, A. Zur Definition der Notenskala. Schweizer Zeitschrift für Psychologie, 1971, 30, 204-218.

FLITNER, A. Das Schulzeugnis im Lichte neuerer Untersuchungen. Zeitschrift für Pädagogik, 1966, 12, 511-538.

FLOOK, A. J. M., & ROBINSON, P. J. Academic performance with and without knowledge of scores on tests of intelligence, aptitude, and personality: A further study. Journal of Educational Psychology, 1972, 63, 123-129.

FLOOK, A. J. M., & SAGGAR, U. Academic performance with, and without, knowledge of scores on tests of intelligence, aptitude, and personality. Journal of Educational Psychology, 1968, 59, 395-401.

FOLDS, J. H., & GAZDA, G. M. A comparison of the effectiveness and efficiency of three methods of test interpretation. Journal of Counseling Psychology, 1966, 13, 318-324.

FRAENKEL, E. Akademisches Prüfungswesen in Deutschland und in den USA. Ein Vergleich. Jahrbuch für Amerikastudien, 1967, 12, 61-73.

FRANK, J. D. Recent studies of the level of aspiration. Psychological Bulletin, 1941, 38, 218-226.

FRETZ, B. R., & ENGLE, D. A. Changes in self-concept as a function of academic test results. Journal of Educational Research, 1973, 66, 227-229.

FROEHLICH, C. P. Does test taking change self ratings? Californian Journal of Educational Research, 1954, 5, 166-169/175.

FROEHLICH, C. P. A criterion for counseling. Psychological Monographs, 1957, 71, whole no. 444.

GAGE, N. L. Paradigmen für die Erforschung des Lehrens. In F. Weinert (Ed.), Pädagogische Psychologie. Köln: Kiepenheuer & Witsch, 1967.

GAIER, E. L. Student self estimates of final course grades. Journal of Genetic Psychology, 1961, 98, 63-67.

GASCH, B. Erfolg im Psychologiestudium. Meisenheim: Hain, 1971.

GEBHARDT, F. Statistische Programme des DRZ. Darmstadt: Deutsches Rechenzentrum, 1969.

GERARD, H. B. Some determinants of self-evaluation. Journal of Abnormal and Social Psychology, 1961, 62, 288-293.

GERGEN, K. J. The concept of self.
New York: Holt, Rinehart & Winston, 1971.

GERTMANN, D., GOMPERT, G., & WEINGARTEN, A. Erste Ergebnisse einer Fragebogenuntersuchung zur Prüfungsvorbereitung im Fach Psychologie. In B. Eckstein (Ed.), Hochschulprüfungen: Rückmeldung oder Repression? Hamburg: AHD, 1971, Blickpunkt Hochschuldidaktik, Heft 13.

GILINSKY, A. S. Relative self-estimate and the level of aspiration. Journal of Experimental Psychology, 1949, 39, 256-259.

GLOCK, M., & MILLMAN, J. The assignment of school marks. In N. E. Gronlund (Ed.), Readings in measurement and evaluation. New York: Macmillan, 1968.

GOSLIN, D. A. Social impact of testing. In G. H. Bracht et al. (Eds.), Perspectives in educational and psychological measurement. Englewood Cliffs: Prentice-Hall, 1972.

GRAUMANN, C. F. (Ed.) Handbuch der Psychologie. Bd. 7.2. Sozialpsychologie: Forschungsberichte. Göttingen: Hogrefe, 1972.

GROEBEN, N., & WESTMEYER, H. Kriterien psychologischer Forschung. München: Juventa, 1975.

GUILFORD, J. P. Fundamental statistics in psychology and education. Tokyo: McGraw-Hill Kogakusha, 1973.

GUTJAHR, W. Die Messung psychischer Eigenschaften. Berlin: VEB Wissenschaften, 1974.

HAAG, F., KRÜGER, H., SCHWÄRZEL, W., & WILDT, J. (Eds.) Aktionsforschung - Forschungsstrategien, Forschungsfelder und Forschungspläne. München: Juventa, 1972.

HAGEN, E. P., & THORNDIKE, R. L. Evaluation. In C. W. Harris (Ed.), Encyclopedia of educational research. New York: Macmillan, 1960.

HAKSTIAN, A. R. The effect on study methods and test performance of objective and essay examinations. Journal of Educational Research, 1971, 64, 319-324.

HARDEE, M. D. Student records and reports - college and university. In C. W. Harris (Ed.), Encyclopedia of educational research. New York: Macmillan, 1960.

HARTOG, P. J. The purposes of examinations. In The yearbook of education for 1938. London: Evans, 1938.

HARTOG, P. J., & RHODES, E. C. An examination of examinations. London: Macmillan, 1936.

HARVEY, O. J., KELLEY, H. H., & SHAPIRO, M. M. Reactions to unfavorable evaluations of the self made by other persons. Journal of Personality, 1957, 25, 393-411.

HASEMANN, K. Studien- und Prüfungsbedingungen im Fach Psychologie an den Wissenschaftlichen Hochschulen der BRD und in West-Berlin. Psychologische Rundschau, 1970, 21, 69-76. (a)

HASEMANN, K. Kriterien der Hochschulreife. Weinheim: Beltz, 1970. (b)

HATCH, W. R. The examination. Improving College and University Teaching, 1958, 6, 126-136.

HAUSNER, S. Die Durchschnittsbildung bei der Zensurengebung. In Arbeitskreis für Hochschuldidaktik (Ed.), Informationen zur Hochschuldidaktik. Hamburg: AHD, 1973, Heft 5.

HECKHAUSEN, H. Hoffnung und Furcht in der Leistungsmotivation. Meisenheim: Hain, 1963.

HECKHAUSEN, H. Leistungsmotivation. In H. Thomae (Ed.), Handbuch der Psychologie. Bd. 2. Motivation. Göttingen: Hogrefe, 1965.

HECKHAUSEN, H. Die Interaktion der Sozialisationsvariablen in der Genese des Leistungsmotivs. In C. F. Graumann (Ed.), Handbuch der Psychologie. Bd. 7.2. Sozialpsychologie: Forschungsberichte. Göttingen: Hogrefe, 1972.

HENTIG, H. von Magier oder Magister? Über die Einheit der Wissenschaft im Verständigungsprozeß. Stuttgart: Klett, 1972.

HERBST, K. Methoden der Ermittlung subjektiver Zensurenskalen an Gymnasien. Unveröffentlichte Diplomarbeit, Psychologisches Institut der Universität Münster, 1972.

HETZER, H. Schüler und Schulzeugnis. Leipzig: Dürr, 1933.

HILLS, D. A., & WILLIAMS, J. E. Effects of test information upon self-evaluation in brief educational-vocational counseling. Journal of Counseling Psychology, 1965, 12, 275-281.

HILLS, J. R. Consistent college grading standards through equating. Educational and Psychological Measurement, 1972, 32, 137-146.

HIMMELWEIT, H. T. Towards a rationalization of examination procedures. Universities Quarterly, 1966/67, 21, 359-372.

HITPASS, J. Das Studienschicksal des Immatrikulationsjahrgangs 1958. Gütersloh: Bertelsmann, 1967.

HÖHN, E. Der schlechte Schüler. München: Piper, 1967.

HOLMES, D. S. Conscious self-appraisal of achievement motivation: The self-peer rank method revisited. Journal of Consulting and Clinical Psychology, 1971, 36, 23-26.

HOLMES, J. E. The presentation of test information to college freshmen. Journal of Counseling Psychology, 1964, 11, 54-58.

HOLT, R. R. Level of aspiration: Ambition or defense? Journal of Experimental Psychology, 1946, 36, 398-416.

HOLZKAMP, K. Begutachtung als Kommunikation. In K. Holzkamp, A. O. Jäger, & F. Merz, Prognose und Bewährung in der psychologischen Diagnostik. Göttingen: Hogrefe, 1966.

HOLZKAMP, K. Wissenschaft als Handlung. Berlin: de Gruyter, 1968.

HOLZKAMP, K. Kritische Psychologie. Frankfurt: Fischer Taschenbuch Verlag, 1972.

HOLZKAMP-OSTERKAMP, U. Grundlagen der psychologischen Motivationsforschung 1. Frankfurt: Campus, 1975.

HONDRICH, K. O. Demokratisierung und Leistungsgesellschaft. Stuttgart: Kohlhammer, 1972.

HOPF, D. Forschungsstand, Forschungsschwerpunkte und Institutionalisierung der pädagogischen Diagnostik. In H. Roth & D. Friedrich (Eds.), Bildungsforschung, Teil 2. Bd. 51, Gutachten und Studien der Bildungskommission, Deutscher Bildungsrat. Stuttgart: Klett, 1975.

HOPPE, F. Erfolg und Mißerfolg. Psychologische Forschung, 1931, 14, 1-62.

HOWARD, R. C., & BERKOWITZ, L. Reactions to the evaluators of one's own performance. Journal of Personality, 1958, 26, 494-507.

HOYT, D. P. College grades and adult accomplishment. Educational Record, 1966, 47, 70-75.

INGENKAMP, K. Pädagogische Diagnostik. Weinheim: Beltz, 1975.

INSKO, C. A. Theories of attitude change. New York: Appleton-Century-Crofts, 1967.

IRLE, M. Lehrbuch der Sozialpsychologie. Göttingen: Hogrefe, 1975.

IRLE, M., & KROLAGE, J. Kognitive Konsequenzen irrtümlicher Selbsteinschätzung. Zeitschrift für Sozialpsychologie, 1973, 4, 36-50.

IRWIN, F. W. The realism of expectations. Psychological Review, 1944, 51, 120-126.

ISRAEL, J. Self-evaluation and rejection in groups. Uppsala: Almquist & Wiksell, 1956.

ISRAEL, J. Sozialpsychologie. Stuttgart: Hippokrates, 1976.

JARRETT, C. D. Marking and reporting practices in the American secondary school. Peabody Journal of Education, 1963, 41, 36-48.

JOHNSON, D. G. Effect of vocational counseling on self-knowledge. Educational and Psychological Measurement, 1953, 13, 330-338.

JOHNSON, H. H. Some effects of discrepancy level on responses to negative information about one's self. Sociometry, 1966, 29, 52-66.

JOHNSON, R. F., & ABRAHAMSON, S. The effects of grades and examinations on self-directed learning. Journal of Medical Education, 1968, 43, 360-366.

JONES, P. M. The fallacies of the examination system. Universities Quarterly, 1964/65, 19, 240-258.

JONES, S. C. Self- and interpersonal evaluations: Esteem theories versus consistency theories. Psychological Bulletin, 1973, 79, 185-199.

JONES, S. C., & SCHNEIDER, D. J. Certainty of self-appraisal and reactions to evaluations from others. Sociometry, 1968, 31, 395-403.

JUOLA, A. E. Illustrative problems in college-level grading. Personnel and Guidance Journal, 1968, 47, 29-33.

KAMINSKI, G. Ordnungsstrukturen und Ordnungsprozesse. In R. Bergius (Ed.), Handbuch der Psychologie. Bd. 1.2. Lernen und Denken. Göttingen: Hogrefe, 1964.

KAMINSKI, G. Verhaltenstheorie und Verhaltensmodifikation. Stuttgart: Klett, 1970.

KAMPMÜLLER, O. Die Beurteilung des Schülers. Bildung und Erziehung, 1961, 14, 88-101.

KANDEL, I. L. Examinations and their substitutes in the United States. New York: Carnegie Foundation, 1936.

KAPUSTE, H. Prüfungen in einem durchlässigen Bildungssystem. In C.-H. Evers & J. Rau (Eds.), Oberstufenreform und Gesamtschule. Frankfurt: Diesterweg, 1970.

KARLINS, M., KAPLAN, M., & STUART, W. Academic attitudes and performance as a function of differential grading systems: An evaluation of Princeton's pass-fail system. Journal of Experimental Education, 1969, 37, 38-50.

KARSDORF, G. Zur erzieherischen Funktion der objektiven Zensierung. Pädagogik, 1961, 16, 530-533.

KATH, G., OEHLER, C., & REICHWEIN, R. Studienweg und Studienerfolg. Berlin: Institut für Bildungsforschung in der Max-Planck-Gesellschaft, 1966, Studien und Berichte. Bd. 6.

KEEFER, R. E. Characteristics of students who make accurate and inaccurate self-predictions of college achievement. Journal of Educational Research, 1971, 64, 401-404.

KEIL, W., & PIONTKOWSKI, U. Strukturen und Prozesse im Hochschulunterricht. Weinheim: Beltz, 1973.

KELLEY, P. R. Jr., MATTHEWS, J. H., & SCHUMACHER, C. F. Analysis of the oral examination of the American Board of Anesthesiology. Journal of Medical Education, 1971, 46, 982-988.

KELLEY, T. L. The use of literal grades. Journal of Educational Psychology, 1950, 41, 488-492.

KENNEDY, W. A., & WILLCUTT, H. C. Praise and blame as incentives. Psychological Bulletin, 1964, 62, 323-332.

KERLINGER, F. N. Grundlagen der Sozialwissenschaften. Bd. 1. Weinheim: Beltz, 1975.

KHANA, P. J., & KHANA, J. L. Current status of the comprehensive examination. American Psychologist, 1972, 27, 761-765.

KIRKLAND, M. C. The effects of tests on students and schools. Review of Educational Research, 1972, 41, 303-350.

KLAFKI, W. Sinn und Unsinn des Leistungsprinzips in der Erziehung. In Sinn und Unsinn des Leistungsprinzips. München: Deutscher Taschenbuch Verlag, 1974.

KNOCHE, W. Jungen, Mädchen, Lehrer und Schulen im Zensurenvergleich. Weinheim: Beltz, 1969.

KOPFMANN, G. Zusammenhang zwischen Abiturnoten von Physikstudenten und ihrer Leistung an der TH München. Zeitschrift für experimentelle und angewandte Psychologie, 1969, 16, 434-451.

KRATHWOHL, D. R., BLOOM, B. S., & MASIA, B. B. Taxonomy of educational objectives. The classification of educational goals. Handbook II: Affective domain. New York: McKay, 1964.

KRISTOF, W. Untersuchungen zur Theorie psychologischen Messens. Meisenheim: Hain, 1969.

KRUMM, V. Der Zusammenhang von Diplomprüfung und Assessorprüfung. Wirtschaft und Erziehung, 1969, 12, 568-573.

KUDER, F. Some principles of interest measurement. Educational and Psychological Measurement, 1970, 30, 205-226.

KURSITZA, H. Die Motivation im Studienprozeß. Hochschulpädagogische Schriftenreihe, 1967, 10, 14-37.

KVALE, S. Prüfung und Herrschaft. Weinheim: Beltz, 1972.

LAUTMANN, R. Gesellschaftliche Mechanismen im Examen. In B. Eckstein (Ed.), Hochschulprüfungen: Rückmeldung oder Repression? Hamburg: AHD, 1971, Blickpunkt Hochschuldidaktik, Heft 13.

LAUWERYS, J. A., & SCANLON, D. G. (Eds.) Examinations. The world yearbook of education 1969. London: Evans, 1969.

LAVIN, W. C. The prediction of academic performance. New York: Russell Sage Foundation, 1965.

LEHMANN, E. Unsere Notenskala. Mathematisches und Pädagogisches. Pädagogische Arbeitsblätter, 1955, 7, 297-304.

LEIBFRIED, S. (Ed.) Wider die Untertanenfabrik. Köln: Pahl-Rugenstein, 1967.

LEONTJEW, A. N. Probleme der Entwicklung des Psychischen. Frankfurt: Athenäum Fischer, 1973.

LERSCH, P. Aufbau der Person. München: Barth, 1962.

LERSCH, P., & THOMAE, H. (Eds.) Handbuch der Psychologie. Bd. 4. Persönlichkeitsforschung und Persönlichkeitstheorie. Göttingen: Hogrefe, 1960.

LEVINE, H. G., & McGUIRE, C. H. The validity and reliability of oral examinations in assessing cognitive skills in medicine. Journal of Educational Measurement, 1970, 7, 63-74.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. S. Level of aspiration. In J. McV. Hunt (Ed.), Personality and behavior disorders. Vol. 1. New York: Ronald Press, 1944.

LIEN, A. J. Measurement and evaluation of learning. Dubuque: Brown, 1967.

LIENERT, G. A. Verteilungsfreie Methoden in der Biostatistik. Bd. I. Meisenheim: Hain, 1973.

LINN, R. L., KLEIN, S. P., & HART, F. M. The nature and correlates of law school essay grades. Educational and Psychological Measurement, 1972, 32, 267-279.

LISTER, J. L., & McKENZIE, D. H. A framework for the improvement of test interpretation in counseling. In N. E. Gronlund (Ed.), Readings in measurement and evaluation. New York: Macmillan, 1968.

LISTER, J. L., & OHLSON, M. M. The improvement of self-understanding through test interpretation. Personnel and Guidance Journal, 1965, 43, 804-810.

LOCH, W. Historische Vorbetrachtungen zu einer Pädagogik der Prüfung. Bildung und Erziehung, 1963, 16, 180-193.

LOCKE, E. A. The relationship of intentions to level of performance. Journal of Applied Psychology, 1966, 50, 60-66.

LOCKE, E. A., & BRYAN, J. F. Grade goals as determinants of academic achievement. Journal of General Psychology, 1968, 79, 217-228.

LOCKE, E. A., CARTLEDGE, N., & KOEPPEL, J. Motivational effects of knowledge of results: A goal-setting phenomenon? Psychological Bulletin, 1968, 70, 474-485.

LORENZ, C. Zehnjahresstatistik des Hochschulbesuchs und der Abschlußprüfungen. II Abschlußprüfungen. Berlin: Verlag für Sozialpolitik, Wirtschaft und Statistik, 1943.

LOWE, C. M. The self-concept: Fact or artifact? Psychological Bulletin, 1961, 58, 325-336.

LUNDBERG, U., & EKMAN, G. Emotional involvement while anticipating an examination: A psychological study. Perceptual and Motor Skills, 1970, 31, 603-609.

McCLELLAND, D. C. Die Leistungsgesellschaft. Stuttgart: Kohlhammer, 1966.

McCLELLAND, D. C. Motivation und Kultur. Bern: Huber, 1967.

McDANIEL, E. D., HALTER, A. N., & HARTFORD, D. L. Grade utility: A new noncognitive factor in academic prediction. Educational and Psychological Measurement, 1961, 21, 621-627.

McGUIRE, C. H. A process approach to the construction and analysis of medical examinations. Journal of Medical Education, 1963, 38, 556-563.

McGUIRE, C. H. The oral examination as a measure of professional competence. Journal of Medical Education, 1966, 41, 267-274.

McINTOSH, D. M., WALKER, D. A., & MACKAY, D. The scaling of teachers' marks and estimates. Edinburgh: Oliver & Boyd, 1962.

McKEACHIE, W. J. Motivation, Lehrmethoden und Lernen in Hochschulen. In F. Weinert (Ed.), Pädagogische Psychologie. Köln: Kiepenheuer & Witsch, 1967.

MADSEN, K. B. Modern theories of motivation. Kopenhagen: Munksgaard, 1974.

MAEHR, M. L., & STALLINGS, W. M. Freedom from external evaluation. Child Development, 1972, 43, 177-185.

MANGOLD, W. Empirische Sozialforschung. Heidelberg: Quelle & Meyer, 1967.

MANNELLO, G. College teaching without grades. Journal of Higher Education, 1964, 35, 328-334.

MARECEK, J., & METTEE, D. R. Avoidance of continued success as a function of self-esteem, level of esteem certainty, and responsibility for success. Journal of Personality and Social Psychology, 1972, 22, 89-107.

MARRIS, P. The experience of higher education. London: Routledge & Kegan, 1964.

MARSHALL, M. S. To grade or not to grade.
Journal of Higher Education, 1952, 23, 264-269/284-285.

MARSHALL, M. S. Teaching without grades.
Corvallis: Oregon State University Press, 1968.

MEANS, R. S., & MEANS, G. H. Achievement as a function of the presence of prior information concerning aptitude.
Journal of Educational Psychology, 1971, 62, 185-187.

MECHANIC, D. Students under stress.
New York: Free Press, 1962.

MEHRENS, W. A. Relation between grade point averages and collegiate course grade distributions.
Journal of Educational Research, 1970, 64, 169-171.

MEICHENBAUM, D. H., & SMART, I. Use of direct expectancy to modify academic performance and attitudes of college students. Journal of Counseling Psychology, 1971, 18, 531-535.

MENCK, P., & THOMA, G. (Eds.) Unterrichtsmethode.
München: Kösel, 1972.

MEYER, G. An experimental study of the old and new types of examinations: I. The effect of the examination set on memory. Journal of Educational Psychology, 1934, 25, 641-661.

MEYER, W.-U. Leistungsmotiv und Ursachenerklärung von Erfolg und Mißerfolg. Stuttgart: Klett, 1973.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. Pläne und Strukturen des Verhaltens. Stuttgart: Klett, 1973.

MILLER, G. E. Evaluation in medical education: A new look.
Journal of Medical Education, 1964, 39, 289-297.

MILLIS, J. S. The role of examinations in the attainment of educational goals. In J. P. Hubbard (Ed.), Examinations and their place in medical education and educational research. Journal of Medical Education, 1966, 41 (special issue), 1-7.

MITSCHERLICH, A. Leistungsverfall - Leistungsverweigerung.
Deutsche Universitätszeitung/Hochschul-Dienst, 1972, [27], 576-581.

MOELLER, M. L. Untersuchungen zur Psychodynamik der neurotischen Prüfungsangst. Dissertation, Berlin, 1967.

MOSCOVICI, S. Der Mensch in der Interaktion: passive Reaktionsmaschine oder wahrnehmungsaktives Subjekt? In S. Moscovici (Ed.), Forschungsgebiete der Sozialpsychologie 1. Frankfurt: Athenäum Fischer, 1975.

MÜLLER, P. Zur Prüfungskritik. In S. Leibfried (Ed.), Wider die Untertanenfabrik. Köln: Pahl-Rugenstein, 1967.

MÜLLER, J. Die juristischen Prüfungen im Deutschen Reich seit 1900. Berlin: Struppe & Winckler, 1932.

MÜLLER, P. H. Lexikon der Stochastik.
Darmstadt: Wissenschaftliche Buchgesellschaft, 1975.

MURSTEIN, B. I. The relationship of grade expectations and grades believed to be deserved to actual grades received. Journal of Experimental Education, 1965, 33, 357-362.

NEUBAUER, W. F. Selbstkonzept und Identität im Kindes- und Jugendalter. München: Reinhardt, 1976.

NIE, N. H., HULL, H. C., JENKINS, J. G., STEINBRENNER, K., & BENT, D. H. SPSS. Statistical package for the social sciences. New York: McGraw-Hill, 1975.

NITSCH, W., GERHARDT, U., OFFE, C., & PREUSS, U. K. Hochschule in der Demokratie. Neuwied: Luchterhand, 1965.

NUNNALLY, J. C. Educational measurement and evaluation. New York: McGraw-Hill, 1972.

OFFE, C. Leistungsprinzip und industrielle Arbeit. Frankfurt: Europäische Verlagsanstalt, 1970.

OPGENOORTH, W. P. Prädiktoren für den Erfolg in einem Statistik-Kurs. Psychologie und Praxis, 1972, 16, 85-89.

OPPENHEIM, A. N., JAHODA, M., & JAMES, R. L. Assumptions underlying the use of university examinations. Universities Quarterly, 1966/67, 21, 341-351.

ORLIK, P. Ein Beitrag zu den Problemen der Metrik und der diagnostischen Valenz schulischer Leistungsbeurteilungen. Zeitschrift für experimentelle und angewandte Psychologie, 1961, 8, 400-408.

OTTWASKA, G. Studienbedingungen - Prüfungsleistungen - Berufserfolg. Hamburg: AHD, 1971, Blickpunkt Hochschuldidaktik, Heft 15.

OWENS, T. R. Predicting academic success in master's degree programs in education. Journal of Educational Research, 1966, 60, 124-126.

PAGE, E. B. Teacher comments and student performance: A seventy-four classroom experiment in school motivation. Journal of Educational Psychology, 1958, 49, 173-181.

PAULSEN, F. Die deutschen Universitäten und das Universitätsstudium. Berlin: Asher, 1902.

PETZEL, T. P. Approval motive and self-estimates of academic performance. Journal of Consulting and Clinical Psychology, 1972, 39, 199-201.

PFANZAGL, J. Die axiomatischen Grundlagen einer allgemeinen Theorie des Messens. Würzburg: Physica, 1959.

PICKUP, A. J., & ANTHONY, W. S. Teachers' marks and pupils' expectations: The short time effects of discrepancies upon classroom performance in secondary schools. British Journal of Educational Psychology, 1968, 38, 302-309.

PIEHL, J. Angst und Konflikt. Experimentelle Untersuchungen an Examenskandidaten. Dissertation. Göttingen, 1972.

PLEASE, N. W. Estimation of the proportion of examination candidates who are wrongly graded. British Journal of Mathematical Psychology, 1971, 24, 230-238.

POKORNY, A. D., & FRAZIER, S. H. An evaluation of oral examinations. Journal of Medical Education, 1966,41, 28-40.

POPHAM, W. J., & HUSEK, T. R. Implications of criterion-referenced measurement. Journal of Educational Measurement, 1969, 6, 1-9.

PORTELE, G. Lernen und Motivation. Weinheim: Beltz, 1975.

PRAHL, H.-W. Hochschulprüfungen - Sinn oder Unsinn. München: Kösel, 1976.

PREISS, H. Zur Frage der objektiven Leistungsbeurteilung bei akademischen Prüfungen. Wissenschaftliche Zeitschrift der Friedrich-Schiller-Universität Jena, Gesellschafts- und Sprachwissenschaftliche Reihe, 1954/55, 4, 569-574.

RADLOFF, R. Affiliation and social comparison. In E. F. Borgatta & W. W. Lambert (Eds.), Handbook of personality theory and research. Chicago: Rand McNally, 1968.

RAIMY, V. C. Self reference in counseling interviews. Journal of Consulting Psychology, 1948, 12, 153-163.

RAPH, J. B., GOLDBERG, M. L., & PASSOW, A. H. Bright underachievers. Studies of scholastic underachievement among intellectually superior high school students. New York: Teachers College Press, 1966.

RAYNOR, J. O. Future orientation and motivation of immediate activity. An elaboration of the theory of achievement motivation. Vervielfältigtes Manuskript. Buffalo: New York State University, 1968.

RAYNOR, J. O. Relationships between achievement-related motives, future orientation, and academic performance. Journal of Personality and Social Psychology, 1970, 15, 28-33.

REESE, J. Auf der Schwelle zwischen Universität und Berufswahl. Frankfurter Hefte, 1970, 25,117-128.

RICHARDS, J. M. Jr. A factor analytic study of the self-ratings of college freshmen. Educational and Psychological Measurement, 1966, 26, 861-870.

RINSLAND, H. D. Constructing tests and grading in elementary and high school subjects. London: Harrap, 1938.

ROBERTSON, M. H. Test scores and self-estimates of two curricular groups. Personnel and Guidance Journal, 1960, 38, 746-750.

ROSENFELD, G. Theorie und Praxis der Lernmotivation. Berlin: VEB Wissenschaften, 1966.

ROSS, C. C. The influence upon achievement of a knowledge of progress. Journal of Educational Psychology, 1933, 24, 609-619.

ROSS, C. C. Should low-ranking college freshmen be told their scores on intelligence tests? School and Society, 1938, 47, 678-680.

ROSS, C. C., & STANLEY, J. C. Measurements in today's schools. New York: Prentice-Hall, 1954.

ROTH, L. (Ed.) Handlexikon zur Erziehungswissenschaft. München: Ehrenwirth, 1976.

RUSAM, A. Erfolg und Versagen im Hochschulexamen. Dissertation, Saarbrücken, 1972.

RUSSELL, D. H. What does research say about self-evaluation? Journal of Educational Research, 1953, 46, 561-573.

SADER, M. Prüfung als Studiensteuerung. Die Zeit, 1968(47), 20-21. (auch unter dem Titel: Prüfungen als Studiensteuerung. In B. Eckstein (Ed.), Hochschulprüfungen: Rückmeldung oder Repression? Hamburg: AHD, 1971, Blickpunkt Hochschuldidaktik, Heft 13.)

SADER, M., CLEMENS-LODDE, B., & WEINGARTEN, A. Kleine Fibel zum Hochschulunterricht. München: Beck, 1970.

SALZMANN, C. Pädagogische Überlegungen zum Problem des Berechtigungswesens. Pädagogische Rundschau, 1966, 20, 134-151.

SANFORD, N. Whatever happened to action research? Journal of Social Issues, 1970, 26, 3-23.

SASSENRATH, J. M., YONGE, G. D., & SHRABLE, K. Immediate and delayed feedback on examinations and immediate and delayed retention. Californian Journal of Educational Research, 1968, 19, 226-231.

SCHALLBERGER, U. Studienverlauf und Studienerfolg. Basel: Beltz, 1974.

SCHEER, J. W., & ZENZ, H. Studenten in der Prüfung. Frankfurt: Aspekte, 1973.

SCHEUCH, E. K. Das Interview in der Sozialforschung. In R. König (Ed.), Handbuch der empirischen Sozialforschung. Bd. 2. Grundlegende Methoden und Techniken, Erster Teil. München: Deutscher Taschenbuch Verlag, 1973.

SCHLIEHE, E., & SEEGER, A. Die verbale Interaktion in mündlichen Hochschulprüfungen. Unveröffentlichte Diplomarbeit, Psychologisches Institut der Universität Münster, 1974.

SCHMIDBAUER, M. Studienorientierung, Studienverhalten und Studienerfolg. Dissertation, München, 1968.

SCHMIDT, H.-D. Leistungschance, Erfolgserwartung und Entscheidung. Berlin: VEB Wissenschaften, 1966.

SCHMIDT, H.-D., WIESENBERG, M., & SCHOLZ, O. B. Über Bedingungen der Veränderung von Nutzenskalen der schulischen Leistungsbewertung. Zeitschrift für Psychologie, 1971, 179, 281-299.

SCHNEIDER, D. J. Tactical self-presentation after success and failure. Journal of Personality and Social Psychology, 1969, 13, 262-268.

SCHREINER, G. Sinn und Unsinn der schulischen Leistungsbeurteilung. Die Deutsche Schule, 1970, 62, 226-237.

SCHREINER, G. Gegen eine verdinglichende Leistungsbeurteilung. Westermanns Pädagogische Beiträge, 1972, 24, 155-159.

SCHÜTZ, M., SKOWRONEK, H., & THIEME, W. (Eds.) Prüfungen als hochschuldidaktisches Problem. Hamburg: AHD, 1969, Blickpunkt Hochschuldidaktik, Heft 1.

SCHUMANN, K. F., & CLAUS, H. J. Prognose des Studienerfolgs. Hamburg: AHD, 1970, Blickpunkt Hochschuldidaktik, Heft 10.

SCHWARZ, E. Experimentelle und quasi-experimentelle Anordnungen in der Unterrichtsforschung. In K. Ingenkamp & E. Parey (Eds.), Handbuch der Unterrichtsforschung, Teil I. Weinheim: Beltz, 1970.

SCIORTINO, R. Factorial study of intellective self-ratings by male and female subjects. Journal of Psychology, 1969, 73, 9-18.

SHEN, E. The validity of self-estimate. Journal of Educational Psychology, 1925, 16, 104-107.

SHERWOOD, J. J. Increased self-evaluation as a function of ambigous evaluation by referent others. Sociometry, 1967, 30, 404-409.

SHRAUGHER, J. S., & ROSENBERG, S. E. Self-esteem and the effects of success and failure feedback on performance. Journal of Personality, 1970, 38, 404-417.

SIEGEL, D. The problem of examination and evaluation in American schools. Internationale Zeitschrift für Erziehungswissenschaft, 1958, 4, 291-305.

SIEGEL, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

SIEGEL, S. Level of aspiration and decision making. Psychological Review, 1957, 64, 253-262.

SILVERMAN, I. Self-esteem and differential responsiveness to success and failure. Journal of Abnormal and Social Psychology, 1964, 69, 115-119.

SIMON, J. G., & FEATHER, N. T. Causal attributions for success and failure at university examinations. Journal of Educational Psychology, 1973, 64, 46-56.

SIXTL, F. Meßmethoden der Psychologie. Weinheim: Beltz, 1967.

SKOLNICK, P. Reactions to personal evaluations: A failure to replicate. Journal of Personality and Social Psychology, 1971, 18, 62-67.

SMITH, A. Z., & DOBBIN, J. E. Marks and marking systems. In C. W. Harris (Ed.), Encyclopedia of educational research. New York: Macmillan, 1960.

SOLWAY, K. S., & FEHR, R. C. Situational variables and self-acceptance. Psychological Reports, 1969, 25, 253-254.

SPAIGHTS, E. Accuracy of self-estimation of junior high school students. Journal of Educational Research, 1965, 58, 416-419.

SPINDLER, D. (Ed.) Hochschuldidaktik. 25 Dokumente zur Hochschul- und Studienreform. Bonn: Verlag Studentenschaft, 1968.

STAHMANN, R. F. Predicting graduation major field from freshman entrance data. Journal of Counseling Psychology, 1969, 16, 109-113.

STALLINGS, W. M., WOLFF, J. L., & MAEHR, M. L. Fear of failure and the pass-fail grading option. Journal of Experimental Education, 1969, 38, 87-91.

STANLEY, J. C., & HOPKINS, K. D. Educational and psychological measurement and evaluation. Englewood Cliffs: Prentice-Hall, 1972.

STONE, L. A. Academic grades: Their subjective scale structure based on college student impressions. Psychological Reports, 1969, 25, 439-442.

STONE, L. A., & SINNETT, E. R. Academic grades: Their rational and empirical scale structure. Psychological Reports, 1968, 22, 681-686.

STOTLAND, E., THORLEY, S., THOMAS, E., COHEN, A. R., & ZANDER, A. The effects of group expectations and self-esteem upon self-evaluation. Journal of Abnormal and Social Psychology, 1957, 54, 55-63.

STRITTMATTER, P. (Ed.) Lernzielorientierte Leistungsmessung. Weinheim: Beltz, 1973.

SUMNER, F. C. Marks as estimated by students. Education, 1932, 52, 429.

TAYLOR, J. B., PTALEK, M., CARITHERS, M., GRIFFIN, C., & COYNE, L. Rating scales as measures of clinical judgement III: Judgements of the self on personality inventory scales and direct ratings. Educational and Psychological Measurement, 1972, 32, 543-557.

TEDESCHI, J. T. (Ed.) The social influence processes. Chicago: Aldine, 1972.

THISTLETHWAITE, D. L. Effects of social recognition upon the educational motivation of talented youth. Journal of Educational Psychology, 1959, 50, 111-116.

THOMAE, H. (Ed.) Handbuch der Psychologie. Bd. 2. Motivation. Göttingen: Hogrefe, 1965.

THORNDIKE, R. L. Marks and marking systems. In G. H. Bracht et al. (Eds.), Perspectives in educational and psychological measurement. Englewood Cliffs: Prentice-Hall, 1972.

THORNDIKE, R. L., & HAGEN, E. P. Measurement and evaluation in psychology and education. New York: Wiley, 1969.

TIMMERBRINK, H. Einstellung zur Prüfung. Unveröffentlichte Diplomarbeit, Psychologisches Institut der Universität Münster, 1971.

TORBERT, W. R. Learning from experience. Toward consciousness. New York: Columbia University Press, 1973.

TORRANCE, E. P. Some practical uses of a knowledge of self-concepts in counseling and guidance. Educational and Psychological Measurement, 1954, 14, 120-127.

TRAVERS, R. M., & GRONLUND, N. E. The meaning of marks. Journal of Higher Education, 1950,21, 369-374.

TRIMBLE, O. C. The oral examination: Its validity and reliability. School and Society, 1934, 39, 550-552.

TYLER, L. E. The work of the counselor. New York: Appleton-Century-Crofts, 1969.

TYLER, R. W. General statement on evaluation. Journal of Educational Research, 1942, 35, 492-501.

ÜBERLA, K. Faktorenanalyse. Berlin: Springer, 1968.

VALLANCE, T. R. A comparison of essay and objective examinations as learning experiences. Journal of Educational Research, 1947, 41, 279-288.

VDS (Verband Deutscher Studentenschaften) Kampfpositionen einer demokratischen Hochschulreform. Deutsche Universitätszeitung /Hochschul-Dienst, 1973, [28], 445-447.

VELDMAN, D. J., & PARKER, G. V. C. Adjective rating scales for self description. Multivariate Behavioral Research, 1970, 5, 295-302.

WALKER, J. L. Four methods of interpreting test scores compared. Personnel and Guidance Journal, 1965, 44, 402-409.

WALL, W. D. Examinations and their bearing on education. Internationale Zeitschrift für Erziehungswissenschaft, 1958, 4, 257-274.

WEINER, B. Theorien der Motivation. Stuttgart: Klett, 1976.

WEINER, B., HECKHAUSEN, H., MEYER, W.-U., & COOK, R. E. Causal ascriptions and achievement behavior: A conceptual analysis of effort and reanalysis of locus of control. Journal of Personality and Social Psychology, 1972, 21, 239-248.

WEINERT, F. Einführung in das Problemgebiet der pädagogischen Psychologie des schulischen Lernens. In F. Weinert (Ed.), Pädagogische Psychologie. Köln: Kiepenheuer & Witsch, 1967.

WEIS, H., & SHERMAN, J. Internal-external control as a predictor of task effort and satisfaction subsequent to failure. Journal of Applied Psychology, 1973, 57, 132-136.

WENKE, H. Akademisches Prüfungswesen in Deutschland: Tatbestände und Probleme. Jahrbuch für Amerikastudien, 1967, 12, 48-60.

WESTBROOK, B. W. The effect of test reporting on self-estimates of scholastic ability and on level of occupational aspiration. Journal of Educational Research, 1967, 60, 387-390.

WILHELMS, F. T. (Ed.) Evaluation as feedback and guide. Washington: National Education Association, 1967.

WILLIAMS, J. E. Changes in self and other perceptions following brief educational-vocational counseling. Journal of Counseling Psychology, 1962, 9, 18-30.

WINER, B. J. Statistical principles in experimental design. Tokyo: McGraw-Hill Kogakusha, 1971.

WINGO, G. M. Methods of teaching (appraising student progress). In C. W. Harris (Ed.), Encyclopedia of educational research. New York: Macmillan, 1960.

WOLFE, R. N. Perceived locus of control and prediction of own academic performance. Journal of Consulting and Clinical Psychology, 1972, 38, 80-83.

WORELL, L. Level of aspiration and academic success. Journal of Educational Psychology, 1959, 50, 47-54.

WOTHKE, W. Untersuchung zur Notengebung bei den Prüfungen am Psychologischen Institut in Münster. Unveröffentlichte Diplomarbeit, Psychologisches Institut der Universität Münster, 1974.

WRIGHT, E. W. A comparison of individual and multiple counseling for test interpretation interviews. Journal of Counseling Psychology, 1963, 10, 126-134.

WULF, C. (Ed.) Evaluation. München: Piper, 1972.

WYLIE, R. C. The self-concept. Lincoln: University of Nebraska Press, 1961.

WYLIE, R. C. Children's estimates of their schoolwork ability as a function of sex, race, and socioeconomic level. Journal of Personality, 1963, 31, 203-224.

WYLIE, R. C. The present status of self theory. In E. F. Borgatta & W. W. Lambert (Eds.), Handbook of personality theory and research. Chicago: Rand McNally, 1968.

ZIELINSKI, J. Macht und Ohnmacht der Zensuren. Pädagogische Rundschau, 1961, 15, 121-142.

Lebenslauf und Bildungsgang:

Geboren wurde ich am 2.1.1944 als Sohn von Elfriede Gertmann, geborene Born, und Artur Gertmann in Freiburg (Breisgau).

Nach der Übersiedlung nach Vlotho/Weser besuchte ich in den Jahren 1950 bis 1954 die Volksschulen in Vlotho und Wehrendorf, in den Jahren 1954 bis 1960 das Progymnasium in Vlotho, in den Jahren 1960 bis 1963 den mathematisch-naturwissenschaftlichen Zweig des Bessel-Gymnasiums in Minden.

Mein Studium begann ich im Sommersemester 1963 in den Fächern Psychologie, Philosophie und Pädagogik an der Universität Münster. Später trat an die Stelle des Faches Philosophie das Fach Soziologie. Ich bestand im Sommersemester 1966 die Diplom-Vorprüfung für Psychologen. Im Wintersemester 1966/67 besuchte ich für ein Semester die Freie Universität Berlin, kehrte im folgenden Sommersemester an die Universität Münster zurück und bestand dort im Wintersemester 1968 die Diplom-Hauptprüfung für Psychologen.

Darnach studierte ich Pädagogik, wurde für ein halbes Jahr mit einem Doktorandenstipendium gefördert und nahm dann eine Arbeit als Wissenschaftlicher Angestellter in einem Herrn Prof. Dr. Sader von der Stiftung Volkswagenwerk finanzierten Forschungsprojekt mit dem Thema: "Empirische Untersuchung der akademischen Prüfung" auf.

Im November 1973 trat ich eine Stelle als Wissenschaftlicher Assistent m.d.V.b. an der Pädagogischen Hochschule Westfalen-Lippe, Abteilung Münster, an. Ich bin dort gegenwärtig als Wissenschaftliche Hilfskraft tätig.